Structural studies of HIV-1 Vif and its SOCS-box domain

Lu, Zhisheng

Awarding institution:
King's College London

Structural studies of HIV-1 Vif and its SOCS-box domain

Zhisheng LU

Randall Division of Cell and Molecular Biophysics

New Hunt's House, Guy's Campus

King's College London

SE1 1UL, London

August 2013

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of London

ABSTRACT

The human immunodeficiency virus type I (HIV-1) is a retrovirus that damages the human immune system, which is suppressed by a cellular factor APOBEC3G in non-permissive cells. The viral infectivity factor (Vif) can induce a poly-ubiquitination degradation of APOBEC3G to counteract the immune response by forming an E3 ubiquitin complex composed of cellular proteins Elongin B (EloB), Elongin C (EloC), Cullin 5 and Ring-Box protein. In this project, we solved the first structure for the Vif SOCS-box and EloBC complex in solution by Nuclear Magnetic Resonance, which shows that the proline-rich motif in the SOCS-box binds to the EloB carboxyl terminus by backbone interaction, based on weak van der Waals forces. Based on the results from cell assays it shows that the residues on the Vif proline-rich motif play an important role in viral infectivity; the proline-rich motif induced structure of the C-terminal tail has been demonstrated to be critical for EloB to perform further biological function, by a mechanism which allows HIV-1 to evade the immune response. In addition, expression and purification on full-length Vif in the presence of EloBC and core binding factor β (CBFβ) has been performed. These studies showed that soluble Vif in the tetramer is achieved when it is expressed at a low temperature with a low IPTG concentration. Solubility tests indicate that this complex must be kept in high salt concentration in order to prevent self-association. Comparison with previous studies on Vif expression and purification. This protocol enables one to obtain purified Vif from *E. coli* and opens the way to solve the structure by X-ray crystallography and extend our understanding on the Vif-induced APOBEC3G degradation mechanism.

DECLARATION

The work in Chapter 3, 4, 5 and 6 are a part of the NMR project that was firstly done by Dr. Julien Bergeron (2010). In the work detailed in this thesis, Dr. Julien Bergeron performed the backbone assignments (unpublished). All the other NMR work performed by the author of this thesis are based on the backbone assignments. Dr. Julien Bergeron's work have been cited properly throughout the thesis.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors, Dr. Mark R. Sanderson and Professor Michael H. Malim, for their patience, help, advice, critique and influence during my Ph.D. training. Their kind instruction guaranteed the progress of my Ph.D. project. I am also grateful to all members of the Sanderson and Malim laboratories including Dennis Veselkov, Ivan Laponogov and Torsten Schaller, Shetal Arjan, Caroline Goujon, Prabh Phalora, Luis Appolonia, Chad Swanson and Hendrik Huthoff. In particular, I am indebted to Dennis Veselkov and Torsten Schaller for their help with protein preparation and functional studies respectively.

I would also like to thank the members of the Randall Division for their help in my project. Especially, this work would not have been possible without the help of Andrew Atkinson who helped me to collect NMR data and provided me a wealth of technical advice, Luigi Martino for the ITC experiments, Rebecca Beavil who performed the ultracentrifuge experiments and Mitla Gracia-Maya for the AKTA facility in a cold cabinet.

Meanwhile, I would like to thank Alain Oregioni, Geoff Kelly and Tom Frenkiel from the NIMR Biomedical NMR centre, Mill Hill, for the help with NMR data collection, and to thank Yi Yang and Stephen Matthews from Division of Molecular Biosciences, Imperial College for the help with the PRE experiments.

I acknowledge the Gaze restaurant that used to be in the basement of Guy's hospital for providing fish and chips every Friday, as I fell in love with this national dish that is disliked by many Chinese.

I would like to thank King's-China Scholarship Council PhD Scholarship Programme that offered me this opportunity to study at King's College London, working with many great colleagues.

Finally, I would like to dedicate this work to Xiaoxin for her support during my preparation of the thesis and the viva.

Table of Contents:

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

2D	two-dimensional
3D	three-dimensional
A3G	APOBEC3G
AIDS	acquired immunodeficiency syndrome
APOBEC3	cytidine to uridine-editing of apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 family
ATP	adenosine triphosphate
AUC	analytical ultracentrifugation
BC-box	elongin BC-binding box
BLAST	basic local alignment search tool
BMRB	biological magnetic resonance bank
BSA	bovine serum albumin
CBFβ	core binding factor β
CcpNnmr	the Collaborative Computing Project for NMR software
CCR5	CC-chemokine receptor 5
cDNA	complementary DNA
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate hydrate
Co-IP	co-immunoprecipitation
COSY	correlation spectroscopy
CS-Rosetta	Chemical-Shift Rosetta
CSGW	a vector name for CMV/spleen focus forming virus strain P long terminal repeat sequence/enhanced GFP/Woodchuck hepatitis virus posttranscriptional regulatory element
CTL	cytotoxic T lymphocyte
Cul	cullin
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid

dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
DVMK	(EloB) aspartate^{101}-valine-methionine-lysine
E. coli	*Escherichia coli*
EloB	elongin B
EloBC	elongin BC dimer
EloC	elongin C
Env	(HIV) envelope protein
GFP	green fluorescent protein
gp	glycoprotein
HA	hemagglutinin epitope
HAART	reverse transcription in highly active antiretroviral therapy
HADDOCK	High Ambiguity Driven biomolecular DOCKing
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV-1	human immunodeficiency virus type I
HIV-2	human immunodeficiency virus type II
HSQC	heteronuclear single quantum coherence spectroscopy
HTS	high-throughput screening
HWI	Hauptman Woodward Medical Research Institute
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
LB	Luria-Bertani broth
MHz	megaHz
mRNA	messenger RNA
MR	molecular replacement method
MS	mass spectrometry
MW	molecular weight
NEB	New England Biolabs
Nef	the negative factor
Ni-NTA	nickel-nitrilotriacetic acid

NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	Nuclear Overhauser Effect spectroscopy
Opti-MEM	optimized minimum essential media
pCMV	a plasmid with cytomegalovirus promoter
PCR	polymerase chain reaction
PEG	polyethylene glycol
PDB	protein data bank
PEI	polyethylenimine
r.m.s.d.	root-mean-square deviation
Pol	(HIV) polymerase
PPLP	(Vif) proline^{161}-proline-leucine-proline
PRE	paramagnetic relaxation enhancement
qPCR	quantitative PCR
Rbx	Ring-finger protein
RCI	random coil index
RDC	Residual dipolar coupling
Rev	(HIV) reverse transcriptase
RNA	ribonucleic acid
SAMHD1	Sterile alpha motif and histidine/aspartic acid domain–containing protein 1
SET-tag	solubility-enhancement-tag
SOCS	the suppressors of cytokine signaling
SOCS-box	SOCS binding domain
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPARTA	Shifts Predicted from Analogy in Residue type and Torsion Angle software
SRS	synchrotron radiation source
ssDNA	single-strand DNA
TALOS	torsion angles from NMR chemical shifts software
Tat	(HIV) transcriptional transactivators
tCBFβ	truncated CBFβ

TEMED	N,N,N´,N´-tetramethylethylenediamine
TMS	tetramethylsilane
TSA	thermal shift assay
TOCSY	total correlation spectroscopy
ToF-MS	Time-of-Flight mass spectrometry
Ub	ubiquitin
Vif	(HIV) viral infection factor
Vpr	(HIV) viral protein R
Vpu	(HIV) viral protein U
Vpx	(HIV-2) viral protein X
WeNMR	a worldwide e-Infrastructure for NMR
WT	wild type

CHAPTER 1: INTRODUCTION

1.1 Acquired Immunodeficiency Syndrome Introduction

Acquired immunodeficiency syndrome (AIDS) is a chronic disease caused by human immunodeficiency virus type I or type II (HIV-I or HIV-II), causing suffers to experience a decreasing immune response capability so that they become more susceptible to opportunistic infections which threaten their lives. Since the first AIDS case was confirmed and reported (Control 1982), there are tens of millions of people dying of HIV/AIDS. By the end of 2010, approximately 33.3 million people lived with HIV/AIDS, 10% of whom were children below 15 years old. More than 2 million people died of AIDS or relative complications and meanwhile around 3 million people are newly infected every year (WHO 2010). Recently, several groups made huge progress in HIV drug and vaccine development (Wu, Yang *et al*. 2010, Wu, Zhou *et al*. 2011, Archin, Liberty *et al*. 2012), HIV/AIDS, however, is still a global problem that threatens all human beings.

1.2 The human immunodeficiency virus

1.2.1 Viral genome and cycling

HIV belongs to the lentivirus family, one of the classes in the retrovirus category (National Institutes of Health 2002). Its genome is composed of nine genes sorted into two classes with three reading frames (Kuiken, Leitner *et al*. 2008). Genes *gag*, *pol* and *env* encode structural proteins and enzymatic proteins including reverse transcriptase, protease and integrase (Feng, Broder *et al*. 1996, Freed 2001). HIV-1, compared to other primitive retroviruses, contains another six genes called accessory genes namely *vif*, *tat*, *rev*, *vpr*, *vpu* and *nef* that encode essential factors believed to engage in the regulation of viral replication and the antagonizing against host restriction factors, i.e. viral infection factor (Vif) (Sheehy, Gaddis *et al*. 2002,

Wiegand, Doehle *et al*. 2004), transcriptional transactivators (Tat) (Gatignol and Jeang 2000, Hetzer, Dormeyer *et al*. 2005, Hong, Lee *et al*. 2013), reverse transcriptase (Rev) (Dundr, Leno *et al*. 1995, Hope 1999), viral protein R (Vpr) (Goh, Manel *et al*. 2004, Tomasicchio, Avenant *et al*. 2013, Wang, Singh *et al*. 2013), viral protein U (Vpu) (Bour, Schubert *et al*. 1995, Neil, Zang *et al*. 2008, Gautam and Bhattacharya 2013, McNatt, Zang *et al*. 2013) and the negative factor (Nef) (Kirchhoff, Greenough *et al*. 1995, Blagoveshchenskaya, Thomas *et al*. 2002, Foster and Garcia 2008, Arhel and Kirchhoff 2009, Lenassi, Cagney *et al*. 2010) respectively (figure 1.1a). These functional proteins will be described in detail later.

HIV-1 virions infect immune CD4+ T cells by recognizing cellular receptors CD4 and CC-chemokine receptor 5 (CCR5) or CXCR4 on cell surfaces via glycoprotein 120 (gp120) on the viral particle surface (figure 1.1b), followed by penetrating through the cell membrane and uncoating in the cytoplasm (Chan and Kim 1998, Markosyan, Cohen *et al*. 2003, Dimitrov, Louis *et al*. 2005), where the viral ribonucleic acid (RNA) is reverse transcribed into deoxyribonucleic acid (DNA) (Isel, Lanchy *et al*. 1996). In the cell nucleus, a portion of the viral DNA is integrated into the host genome and is copied and transcribed; This is responsible for ongoing viral generation in infected cells, and the rest of the DNA forms a circular DNA called the long terminal repeat DNA (Pauza, Trivedi *et al*. 1994, Gillim-Ross, Cara *et al*. 2005). Generated viral mRNA that is exported into the cytoplasm encodes for structural proteins that package viral RNA and assemble near the cell membrane (Yeager, Wilson-Kubalek *et al*. 1998, Wilk, Gross *et al*. 2001, Briggs, Simon *et al*. 2004), forming new HIV particles. A mature retrovirus is finally released after budding from the cell membrane (Swanstrom and Wills 1997, Pornillos, Garrus *et al*. 2002).

1.2.2 Accessory proteins and the mechanism for evading immune response

HIV-1 develops several methods to suppress various immune defenses via six accessory proteins in the genome as mentioned above. Tat, a 14-kDa protein encoded by two exons, is critical for Tat-mediated viral transactivation. Basically, it forms an adenosine triphosphate (ATP)-dependent RNA-Tat complex with other cellular factors, which performs the transactivitation function, including promoter clearance and elongation (Brigati, Giacca *et al*.

2003, Bannwarth and Gatignol 2005, Hetzer, Dormeyer *et al*. 2005). Rev is a 116-residue nucleolar protein with four functional domains, termed the nuclear localization signal, RNA binding domain, oligomerization domain and nuclear export signal, generated by spliced mRNAs (Feinberg, Jarrett *et al*. 1986, Sodroski, Goh *et al*. 1986, Cochrane, Perkins *et al*. 1990). The nuclear export signal, a leucine-rich motif, is important for Rev function because of the interaction with its nuclear export receptor CRM1, a conserved nuclear receptor from yeast to humans that regulates the export of many proteins (Fischer, Huber *et al*. 1995, Meyer, Meinkoth *et al*. 1996, Nekhai and Jeang 2006). Besides CRM1, a set of cellular factors, such as some members of DEAD-box family, are recruited by Rev to function at different stages as well, which is already detailed in the literature (Reddy, Tang *et al*. 2000, Hofmann, Reichart *et al*. 2001, Fang, Kubota *et al*. 2004, Yedavalli, Neuveut *et al*. 2004, Yu, Sanchez-Velar *et al*. 2005, Joyner, Keuper *et al*. 2013).

Antigen-specific CD8-mediated cytotoxic T lymphocytes (CTL) are important immune response in the host immune system and are able to recognize and kill HIV-infected cells in less than 6 hours (Schmitz, Kuroda *et al*. 1999). HIV-1 Vpr, a small protein with only 96 amino acid residues conserved in most lentiviruses (Tristem, Purvis *et al*. 1998), is able to suppress the host immune activation by arresting cell cycling at G_2/M phase, giving approximately 2 hours for replication (Levy, Refaeli *et al*. 1994, Ayyavoo, Muthumani *et al*. 2002). Heat stress response is another anti-HIV immunity mechanism to protect the host against infection by reducing the Vpr-dependent G2 arrest. Vpr, however, can counteract this effect by down-modulating the production of heat shock proteins (Iordanskiy, Zhao *et al*. 2004, Iordanskiy, Zhao *et al*. 2004, Benko, Liang *et al*. 2007, Wei, Guo *et al*. 2012).

Tetherin, a factor that had unknown function before, is a membrane protein sitting at the cell surface. It is also named CD317, HM1.24 or BST-2, and it is found that Vpu-deleted HIV-1 virions fail to be released in the presence of Tetherin, i.e. they ‘tether’ mature viral particles onto the cell surface. Tetherin usually keeps a low expression level whereas type I interferons will induce high expression *in vivo* (Homann, Smith *et al*. 2011). The depletion of Tetherin abolishes the requirements of Vpu (Neil, Zang *et al*. 2008, Van Damme, Goff *et al*. 2008). Later studies reveal that Vpu-dependent immune neutralization is species-specific (Goffinet, Allespach *et al*. 2009, McNatt, Zang *et al*. 2009). It reduces the level of Tetherin on the cell

surface by launching the degradation of Tetherin via β-TrCP-dependent lysosomal pathway and endosome-dependent proteasomal pathway (Douglas, Viswanathan *et al*. 2009, Mangeat, Gers-Huber *et al*. 2009). Considering that the down-regulation of Tetherin is not always demanded for Vpu-dependent virion release (Dube, Roy *et al*. 2010, Goffinet, Homann *et al*. 2010), this antagonizing-Tetherin mechanism is not the only mechanism that Vpu mediates against the blockage of virion release, which means further research is required to clarify the Vpu-induced immunal evasion. Meanwhile, some groups now propose that the selection pressure on Vpu by Tetherin is critical for the function of Vpu, suggesting a competition between hosts and virus during the evolution (Sauter, Unterweger *et al*. 2012, Douglas, Bai *et al*. 2013).

HIV-1 Nef protein is a 27 kDa myristoylated protein localized on cytoplasmic membranes, with multiple functions that are required for viral recycling in cells. Besides the function of Tetherin antagonizing Vpu (Jia, Serra-Moreno *et al*. 2009, Zhang, Wilson *et al*. 2009, Sauter and Kirchhoff 2011), Nef down-regulates the level of major histocompatibility class I on cell surface so that infected cells can escape the antiviral effect by cytotoxic T lymphocytes cells (Blagoveshchenskaya, Thomas *et al*. 2002, Kasper, Roeth *et al*. 2005). Similarly, removing cellular surface receptors such as CD4, CCR5 and CXCR4 by inducing endocytic or lysosome degradation can help viruses escape CD4+ helper T cells-driven immune responses and prevent signal transduction among infected cells (Aiken, Konner *et al*. 1994, Lundquist, Tobiume *et al*. 2002, Michel, Allespach *et al*. 2005, Sloan, Donahue *et al*. 2010). Immunoglobulin class switching in B cells is critical in immunity against the virus. Nef is found to evade B cells and disrupt the immunoglobulin G and immunoglobulin A switching, which counteracts antibody response that neutralizes infected cells (Qiao, He *et al*. 2006). Another novel function of Nef is to remove a kinase called Lck in the Src family from membrane micro-domains and to disrupt the vesicular transport of Lck, thus tailoring T-cell activation and optimizing virus replication (Pan, Geist *et al*. 2013), which provides new insight on the wide mechanisms of Nef behaviour in cells. More detail on Nef function have been described comprehensively in several reviews (Roeth and Collins 2006, Foster and Garcia 2008, Arhel and Kirchhoff 2009). Sterile alpha motif and histidine/aspartic acid domain–containing protein 1 (SAMHD1) is another protein that affects HIV infection by controlling nucleic acid metabolism, which is antagonized by viral protein x (Vpx) from HIV-2 (Schaller, Goujon *et al*. 2012). The host intrinsic defense by cellular

factors and the corresponding viral evasions are summarized in table 1.1 and figure 1.2.

Table 1.1, Host intrinsic defenses and mechanisms of viral evasion or antagonism

Host defence	Antiviral effect	Viral factors for evasion
APOBEC3G	Induce lethal hypermutations	Vif: Induce ubiquitination of APOBEC3G
	Interfere with elongation process	
	Activate global immune responses	
Trim 5α	Interfere with uncoating	Viral variability
Tetherin	Block viral release	Vpu, Nef: Induce degradation of Tetherin
SAMHD1	Reduce dNTP required for cDNA production	Vpx: Induce proteaseome degradation

1.2.3 Viral protein structural studies

HIV/AIDS studies appeal to many structural biologists' interest as well as to cytology and virology biologists, mainly because of the requirement for the understanding of basic biology from a molecular viewpoint and for the development of new anti-HIV drugs. In the past dozen years, there have been intensive reports on HIV-1 viral particle structures and viral protein structures, compared to other viruses or protein families. Architectures on HIV-1 structural proteins such as Gag and envelope protein (Env), accessory proteins including Vpu, Vpr and Tat, and proteases like transcriptases and integrases have been reported. Some of these structural data have already been applied in vaccine and drug research.

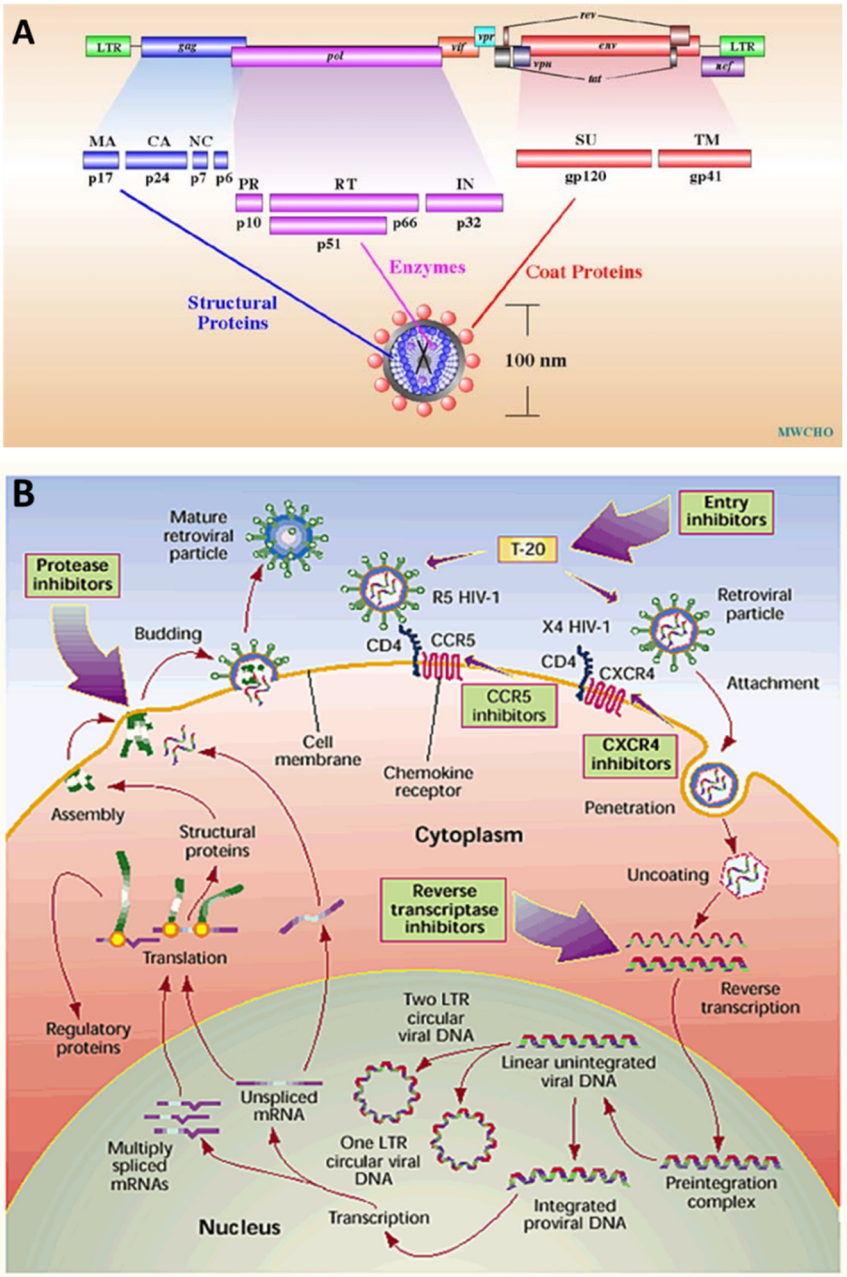


Figure 1.1: (A) The genome and viral proteins of HIV-1. Proteins (middle) translated from the viral genome (top) are assembled into virions. Genes that share the same open reading frame are shown in the different colours. LTR, long terminal repeat. The figure is reproduced from website: http://www.stanford.edu/group/virus/retro/2005gongishmail/HIV.html. (B) The summary of HIV-1 life cycle. Reverse transcription and translation complete in the cytoplasm (red region) and DNA replication takes place in the nucleus (grey region). Relative inhibitors used in each replication step are highlighted in green boxes. Reproduced from reference: HIV-1 entry inhibitors: evading the issue (Michael and Moore 1999).

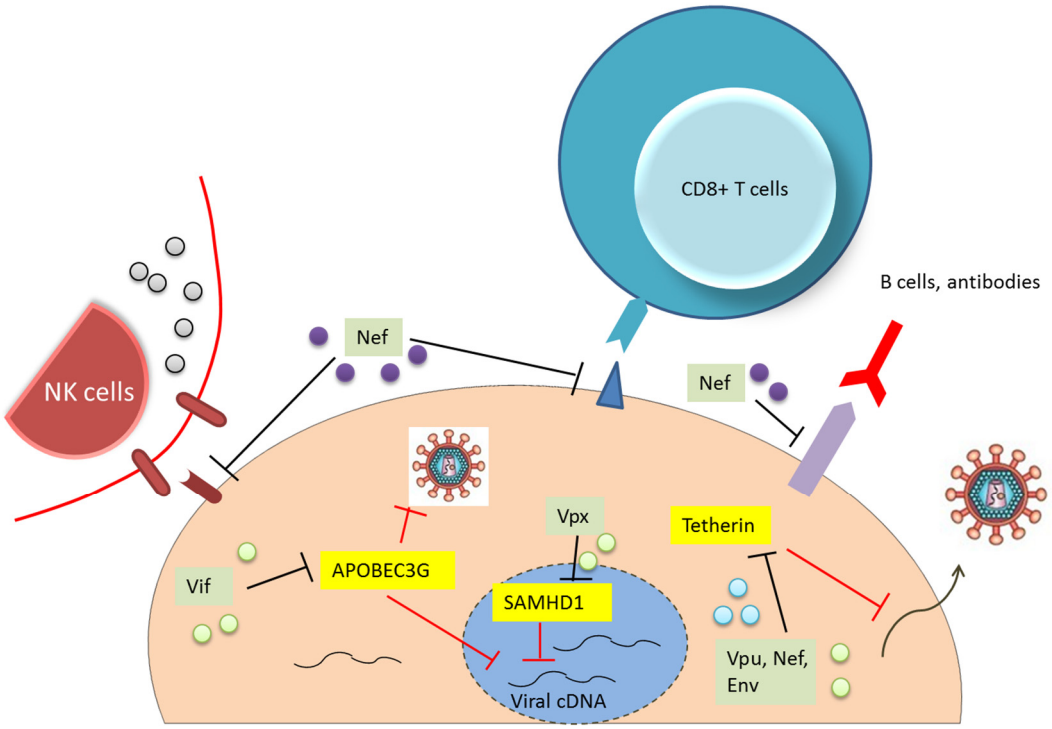


Figure 1.2: The summary of host defense reactions and viral counteracting pathways. Host immune factors are highlighted in yellow and viral functional accessory proteins are highlighted in green. The blue region in the middle stands for nucleus. The competition between host and virus is summarised in table 1.1.

Env spikes consist of a glycoprotein trimer, gp120, linked non-covalently to a transmembrane glycoprotein trimer gp41 on the viral surface. The heterodimer so-formed dominates a cascade of conformational changes that initiates fusion between cell membranes and viral particles (figure 1.3A) (Zhu, Liu *et al*. 2006, Liu, Bartesaghi *et al*. 2008). A cellular surface receptor CD4 engages an interaction with core gp120, followed by a second binding to CCR5 and CXCR4 in the recognition and fusion step (figure 1.3B) (Kwong, Wyatt *et al*. 1998, Rizzuto, Wyatt *et al*. 1998, Pancera, Majeed *et al*. 2010). Some groups reported that this interface can be targeted as an inhibiting site by CD4-like molecules to neutralize virus, and that some positive results have already been achieved with this approach (Wu, Yang *et al*. 2010, Wu, Zhou *et al*. 2011). Gag, also called viral capsid protein, is a structural protein that forms a HIV core shell in which replication enzymes and viral RNA are housed (Briggs, Grunewald *et al*. 2006). Recent studies reveal that the core capsid structure is assembled by hexameric and pentameric Gags as ring structures (figure 1.4) (Pornillos, Ganser-Pornillos *et al*. 2009, Pornillos, Ganser-Pornillos *et al*. 2011) and encouragingly the mature capsid structure was published this year, thus providing a comprehensive knowledge on the HIV-1 capsid (Zhao, Perilla *et al*. 2013).

HIV-1 reverse transcriptase and integrase are both encoded by the *pol* gene. The structure of reverse transcriptase with an inhibitor or in the presence of DNA was reported twenty years ago (Kohlstaedt, Wang *et al*. 1992, Jacobo-Molina, Ding *et al*. 1993). A multi-domain structure of HIV-1 integrase was reported a dozen years ago (Chen, Krucinski *et al*. 2000) whereas useful and universal structural mechanism of integrase were achieved from other integrases in the retrovirus family quite recently because of solubility problem (Hare, Gupta *et al*. 2010, Maertens, Hare *et al*. 2010). HIV-1 reverse transcriptase is composed of a p51 subunit and a p66 subunit containing four folded sub-domains named Fingers, Palm, Thumb and Connection (figure 1.5A). The active site for the excision reaction is located in the p66 Palm subunit and mutates under pressure to acquire drug-resistance to enhance the ATP-regulated excision at the 3´-hydroxyl of the DNA template (Jacobo-Molina, Ding *et al*. 1993, Tu, Das *et al*. 2010). The HIV-1 integrase C-terminus forms a dimer whose catalytic core is exposed inside the dimer interface, serving to bind DNA, bending and orienting it during integration (figure 1.5B) (Eijkelenboom, Sprangers *et al*. 1999, Chen, Krucinski *et al*. 2000). The crystal structure of the

integrase from the prototype foamy virus indicates that the N-terminus of integrase functions to stabilize the tetramer by the N-terminal domain linking to the catalytic core domain although it does not bind directly to DNA (Hare, Gupta *et al*. 2010).

Crystallographic studies on Rev present that Rev behaves as a dimer in solution. The arginine-rich motif is on the top of this V-shape structure to interact with adjacent RNA (figure 1.6A). A secondary binding interface is found to adopt a helical hairpin that is critical in Rev multimerization by hydrophobic interaction, which later binds to host export factor Crm1 (Daugherty, Liu *et al*. 2010, DiMattia, Watts *et al*. 2010). The arginine-rich motif forms a hydrophilic surface that binds to RNA in the Rev polymer. Tat is largely unfolded as shown by nuclear magnetic resonance (NMR) studies when not bound, because the strictly conserved core region is highly flexible in the solution (Bayer, Kraft *et al*. 1995). The crystal structure shows that Tat is complementary to the surface of host protein positive transcription elongation factor by its arginine-rich α-helix (figure 1.6B) (Tahirov, Babayeva *et al*. 2010). The structure of Vpu cytoplasmic domain was solved in micelles by solution NMR (Wittlich, Koenig *et al*. 2009). The dodecylphosphatidylcholine micelles closely mimic the actual condition in cells for transmembrane proteins. It shows that Vpu holds a high degree of flexibility as well as Tat and Vpr (Schuler, Wecker *et al*. 1999) and its helical regions are relevant to membrane attachment. More comprehensive details on HIV-1 protein structures can be found in a recent review (Engelman and Cherepanov 2012).

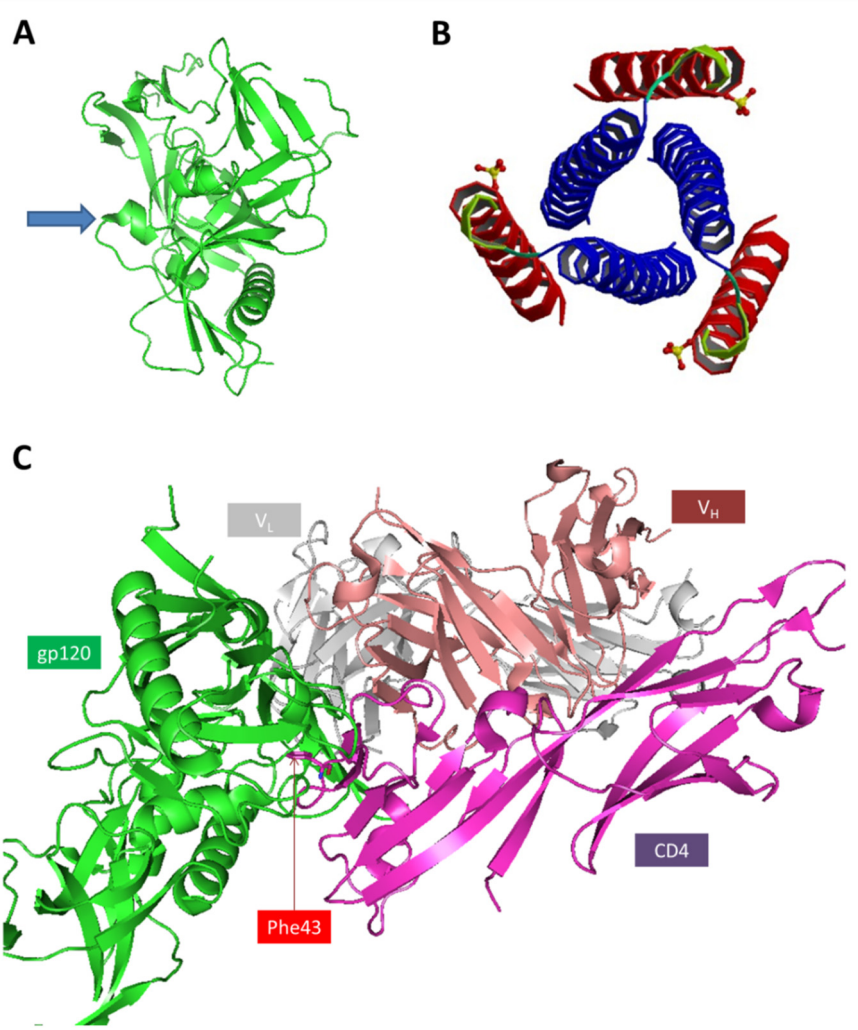


Figure 1.3: The structures of HIV-1 glycoproteins. (A) The structure of HIV-1 gp120 monomer. The left surface indicated by a blue arrow is bound to the broadly neutralizing antibody b12. This structure is regenerated from the crystal structure, PDB accession: 2NY7. (B) The structure of HIV-1 gp41 trimer (top view). A blue α helix and a red one forms a gp41 monomer. The α-helical trimer crosses the cellular membrane and binds to gp120. PDB ID: 1QR8 (Ji, Shu *et al*. 1999). (C) The gp120-CD4 interaction. The structure of HIV-1 gp120 in the complex with cellular surface glycoprotein CD4 and FAB antibody. Phe43 on the receptor forms a hydrophobic cavity inside the complex highlighted in red. (V_L stands for light chain and V_H stands for heavy chain). Reproduced from crystal structure, PDB ID: 3JWD. All reproduced structure images are generated using Pymol.

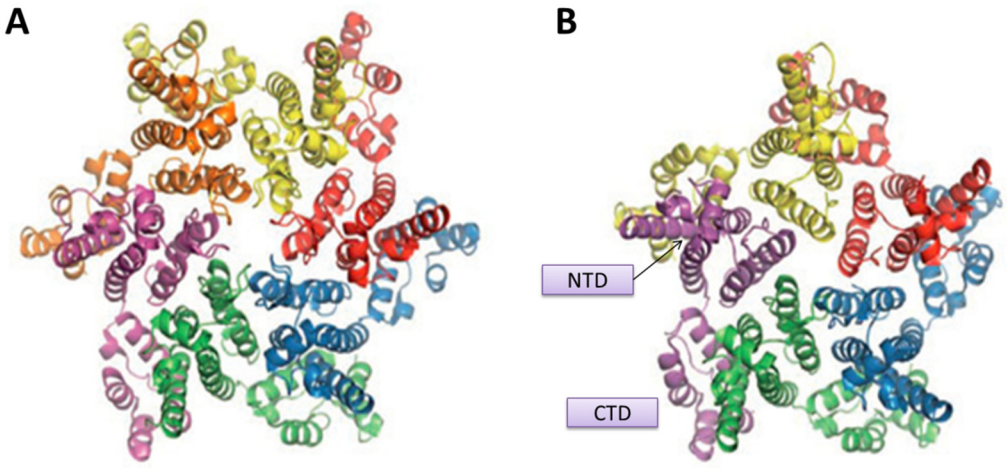


Figure 1.4: The structures of HIV-1 gag. (A) The structure of hexameric HIV-1 capsid protein. (B) The structure of pentameric HIV-1 capsid protein. Each monomer is presented in a different colour. View from the top of the protein. Each subunit is coloured individually. Figures are adapted from: The structural biology of HIV‐1: mechanistic and therapeutic insights (Engelman and Cherepanov 2012).

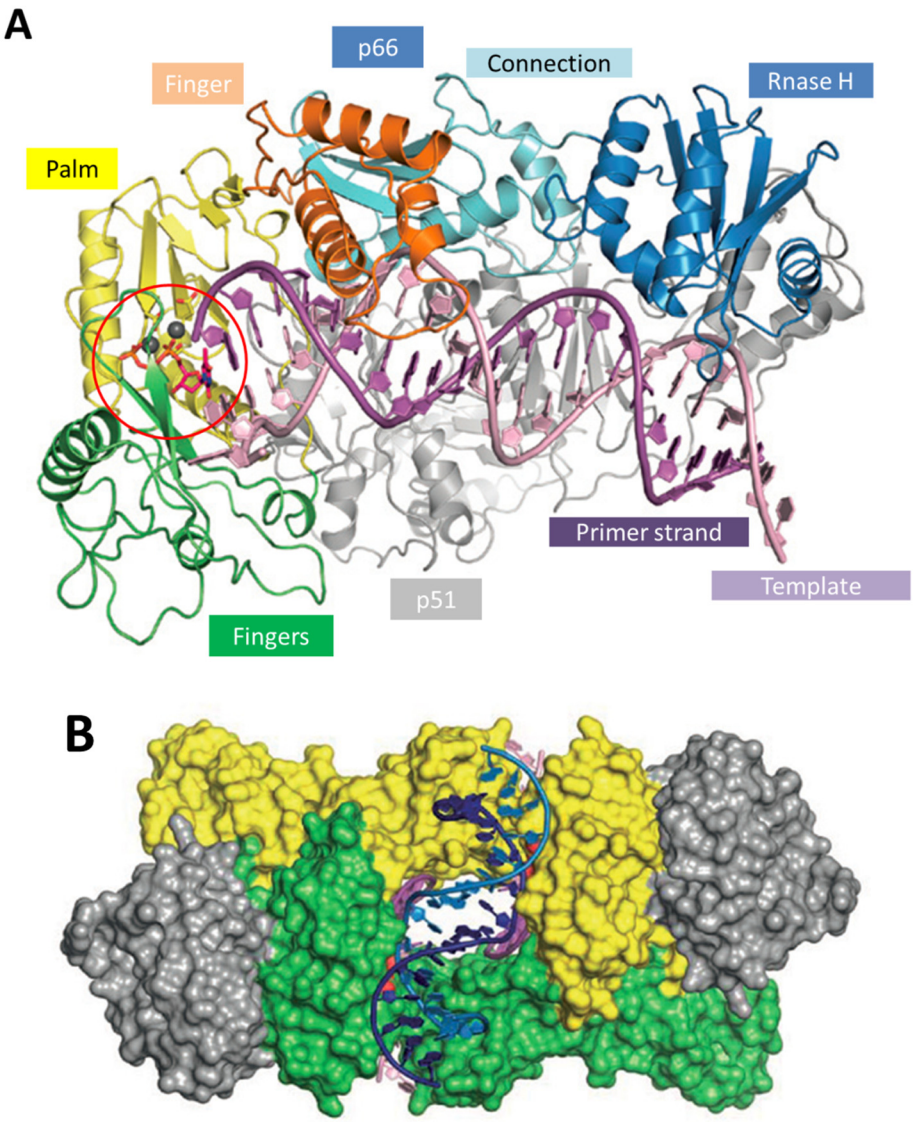


Figure 1.5: HIV-1 polymerase proteins. (A) The structure of HIV-1 reverse transcriptase in the process of DNA genesis. The active site is indicated in red and four functional domains are highlighted in different colours. (B) The general view of the prototype foamy virus integrase structure. The N-terminus (coloured in grey) interacts with the C-terminus from the other subunit. Two C-terminus forms the catalytic core domain. DNA is embedded in the middle of the dimer. Figures are adapted from: The structural biology of HIV‐1: mechanistic and therapeutic insights (Engelman and Cherepanov 2012).

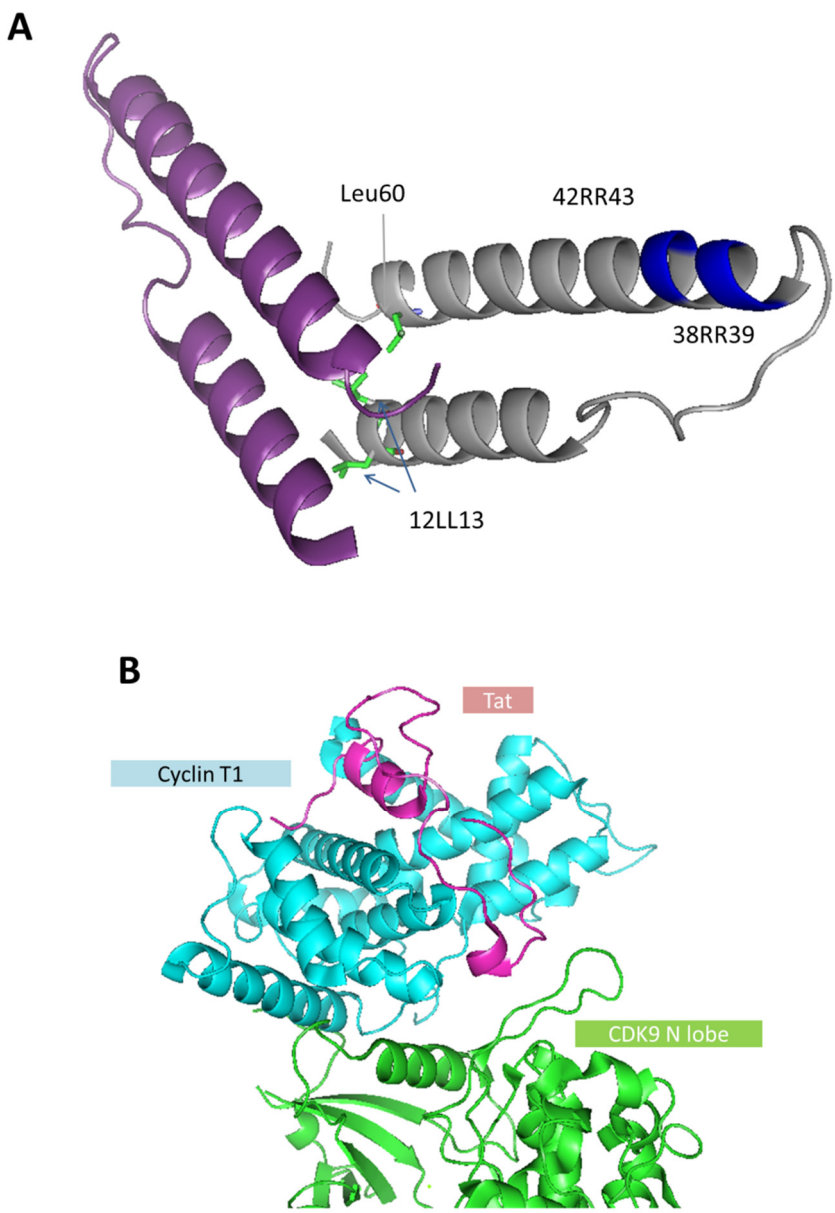


Figure 1.6: HIV-1 accessory protein structures. (A) The dimer of HIV-1 Rev protein. Each monomer is coloured according to chain. The side-chains of three leucines on the interface are drawn in sticks in green. The positions of arginines 38, 39, 42 and 43 is indicated in blue. Reproduced from crystal structure 2X7L. (B) The structure of Tat (the unstructured peptide in pink) observed in crystal structure 3MIA. Only Tat, Cyclin T1 and CDK9-N-lobe are presented here. PDB ID: 3MIA.

1.3 Cellular ubiquitination pathway

Vif, an accessory protein that helps HIV survive in the non-permissive cells, functions by inducing ubiquitination on APOBEC3F and APOBEC3G. Before the review on Vif-APOBEC3G interaction, the general mechanism of ubiquitination pathway will be introduced in this section first.

1.3.1 Overview of ubiquitination mechanism

Ubiquitin (Ub), originally known as ubiquitous immunopoietic polypeptide, is a 76-amino-acid-residue protein covalently attaching to a large range of proteins in cells. The function of ubiquitin, termed ubiquitination, is referred to as a post-translational modification, resulting in protein turn-over and degradation. It is demonstrated that the linkage of ubiquitin to the target protein is via several lysine linkage types (Peng, Schwartz *et al*. 2003, Xu and Peng 2008).

The ubiquitination process consists of three major steps (figure 1.7). In the first step, an E1 ubiquitin-activating enzyme binds to a free ubiquitin using the energy released by ATP. Then the ubiquitin on E1 is transferred to an E2 ubiquitin-conjugating enzyme. In the third step, the E2 domain binds to an ubiquitin and is recognized by E3 ubiquitin-ligase enzyme itself. Binding to the target protein helps to transfer the ubiquitin from E2 to the substrate via two domains, HECT or RING domain (Deshaies and Joazeiro 2009, Hochstrasser 2009, Rotin and Kumar 2009). One of the members in Ring family is known as the Cullin-Ring enzyme, where structural studies showed that it is divided into two subunits. The E3 core is usually composed of Ring-finger protein (Rbx), Cullin protein (Cul), a substrate-recognition portion and other proteins, some structures of which have been published (Zheng, Schulman *et al*. 2002, Duda, Borg *et al*. 2008). As for the Vif-dependent pathway in HIV-1 infected cells, the E3 core complex is composed of proteins Cullin 5 (Cul5), elongin B (EloB), elongin C (EloC) and Rbx (Yu, Yu *et al*. 2003, Kamura, Maenaka *et al*. 2004, Yu, Xiao *et al*. 2004).

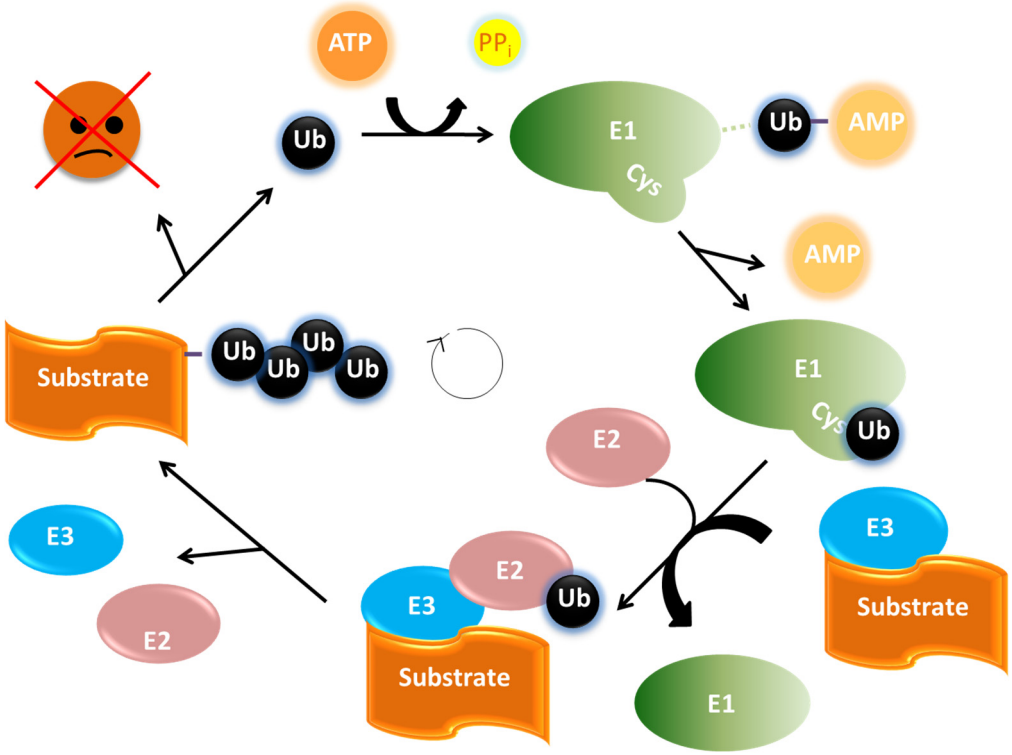


Figure 1.7: Ubiquitination pathway. The ubiquitin molecule is passed onto the substrate by several Elongin molecules. In the first step E1 recruits ubiquitin through its Cys residues, which is supplemented by ATP. The ubiquitin is then passed onto E2 that binds to E3 complex. In the final step the ubiquitin binds to the substrate followed by E2 and E3 dissociate. This cycle repeats several times until the substrate is labeled completely and transferred to the degradation pathway. Reproduced from Mark Hochstrasser's review: Origin and function of ubiquitin-like proteins (Hochstrasser 2009).

1.3.2 The structural study on Elongin B and Elongin C

Elongin BC dimer (EloBC) was originally identified in the Elongin ABC heterotrimer complex in 1995 (Aso, Lane *et al*. 1995). Later it was found that the suppressors of cytokine signaling (SOCS) factors also participate in protein degradation by interacting with EloBC, which reveals that the EloBC protein degradation pathway is connected to other transduction pathways (Zhang, Farley *et al*. 1999). In the ligase complex EloC is responsible for selectively recruiting the Cullin/Rbx complex in comparison to an EloC-like adaptor Skp1 protein that recruits other factors to form the complex (Yan, Kamura *et al*. 2004).

Structural studies reveal that EloB and EloC form a heterodimer via their hydrophilic β-sheet domains and the C-terminal tail on EloC interacts with a conserved domain on other proteins, namely the BC-box because of its binding to EloBC (Bullock, Debreczeni *et al*. 2006, Bullock, Rodriguez *et al*. 2007, Stanley, Ehrlich *et al*. 2008). NMR perturbation shows that both EloB and EloC experience a folding process during the binding to other proteins and EloC is partly unfolded before binding (figure 1.8) (Babon, Sabo *et al*. 2008, Knauth, Cartwright *et al*. 2009, Bergeron, Huthoff *et al*. 2010). It is of note that there are no reports of producing soluble EloB or EloC individually from E. *coli*.

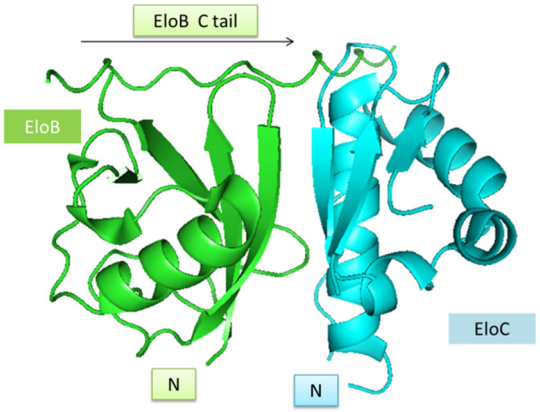


Figure 1.8: The structure of EloBC heterodimer. The direction of the partial flexible EloB C-terminal that is observed by X-ray crystallography is indicated. Reproduced from crystal structure, PDB ID: 3DCG.

1.4 APOBEC3 Family and the Vif-induced degradation of APOBEC3G

1.4.1 Overview of APOBEC3 family

The apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 (APOBEC3, A3) family of cytidine to uridine-editing enzymes contain seven members named from A to H (LaRue, Andresdottir *et al*. 2009) that are constitutively present in cells exhibiting physiological functions such as triggering deamination pathways in antibody gene diversification processes and anti-virus defence (Petersen-Mahrt, Harris *et al*. 2002, Harris, Sheehy *et al*. 2003, Conticello, Thomas *et al*. 2005, Pham, Bransteitter *et al*. 2005, Rosenberg and Papavasiliou 2007, Kohli, Maul *et al*. 2010). All of these genes are located on choromosome 22 (Jarmuz, Chester *et al*. 2002). APOBEC3G (A3G) is a cellular factor with 384 amino acid residues translated from the a3g gene. It is the one that is known to play the most important role in viral suppression, especially in the anti-human immunodeficiency virus type 1 (HIV-1) processes, among the APOBEC3 family members, which have recently been comprehensively described in several reviews (Albin and Harris 2010, Kirchhoff 2010, Wissing, Galloway *et al*. 2010, Imahashi, Nakashima *et al*. 2012).

1.4.2 Functional and structural studies on APOBEC3G

The function of APOBEC3G to inhibit HIV was first identified in 2002 by Michael Malim's group when this factor was named CEM15 at the time. They tested the expression level of CEM15 in permissive and non-permissive cells during the infection with or without Vif. It was found that HIV was suppressed by A3G in the absence of Vif, whereas this inhibition activity of A3G was overcome by Vif (Madani and Kabat 1998, Simon, Gaddis *et al*. 1998, Sheehy, Gaddis *et al*. 2002). A later study, however, showed that even Vif-deficiency HIV can still survive in the APOBEC3G-expressing cells, indicating that other mechanisms may exist to counteract the function of APOBEC3G (Hache, Shindo *et al*. 2008).

APOBEC3G suppresses Vif-deficiency HIV-1 in non-permissive cells by being packaged in mature viral particles into next generation cells. The reverse transcription complexes of HIV-

1 are not affected by A3G packaging, which means this mechanism is independent of HIV reverse transcription. After the virus has fused and decoated in target cells, A3G induces numerous mutations on retroviruses, especially cytidine to uridine mutations in the single-strand viral DNA by deamination mechanism in host cells (Harris, Bishop *et al*. 2003, Lecossier, Bouchonnet *et al*. 2003, Mangeat, Turelli *et al*. 2003). The amount of A3G that is delivered into peripheral blood mononuclear cells is around seven molecules, suggesting that only a few A3G molecules are required for inhibiting HIV (Xu, Chertova *et al*. 2007). In another investigation, it was shown that that single packaged A3G molecule can induce the mutation as well by establishing low-expression cell lines (Browne, Allers *et al*. 2009). This molecular mechanism takes place by directly attacking position 4 of the pyrimidine ring in cytidine where zinc is required to bind the pyrimidine ring, inducing a substitution of the 4′ amino-group with a carbonyl-group (figure 1.9A) (Neuberger, Harris *et al*. 2003).

Besides viral DNA degradation, APOBEC3G, as well as APOBEC3F, is able to suppress the virus by impeding viral DNA reverse transcription, specifically, by inhibiting reverse transcriptase-catalyzed DNA elongation reactions in target cells in the absence of hypermutation (Holmes, Koning *et al*. 2007, Iwatani, Chan *et al*. 2007, Bishop, Verma *et al*. 2008). Recently a novel study on APOBEC3G showed that the expression of APOBEC3G was not only acting as a viral defense, but also as an activator of the immune system (Casartelli, Guivel-Benhassine *et al*. 2010). In the present of APOBEC3G, CD8+ cytotoxic T lymphocytes were significantly activated. The function of APOBEC3G is considered much more complex than expected because it is involved in the host global immune system rather than just the stimulation of mutations.

Since the first time that APOBEC3G (CEM15) was found to inhibit HIV-1 in the absence of Vif (Sheehy, Gaddis *et al*. 2002), the structure solution of APOBEC3G and its related complex naturally became a critical target due to the possibility for treating HIV/AIDS. In 2008 the C-terminal 187-amino-acid-residue structure was solved both in solution by NMR by Chen's group (Chen, Harjes *et al*. 2008) and by X-ray crystallography Holden's group respectively (Holden, Prochnow *et al*. 2008).

The NMR solution structure suggests that the C-terminus of A3G (198-384) containing the deamination activity is composed of five α-helices and five β-strands (figure 1.9B) (Chen,

Harjes *et al*. 2008). The crystal structure is composed of six α-helices, one of which is not defined at the N-terminus in the NMR structure, and five β-strands. Notably, the β2 strand in the crystal structure is an intact β strand whereas in the solution structure it is short. This difference may be caused by the mutation in the NMR structure, implying that five β-sheets are a common structure in the APOBEC family (Holden, Prochnow *et al*. 2008). Alignment of APOBEC2 and APOBEC3G crystal structure shows that the core structures of the APOBEC family is highly conserved. However, APOBEC3G has a longer active centre loops. For example, loop 3 in the crystal structure is a deaminase domain (figure 1.9C), where mutation at R215 will abolish the deamination activity because R215 is responsible for an interaction with R313 and W285 by hydrogen bonds (figure 1.9D). Besides, R256 and R264 in the loop form a strong salt bridge that stabilizes the loop. The single-strand DNA (ssDNA) acts as a substrate and is embedded in the groove on the surface of A3G starting from loop1 to loop3, where the target base of the ssDNA is located in the active site so that the Zn^{2+}-induced deamination can take place. These newly discovered features revealed by structural biological methods explain the deamination mechanism and provide a knowledge that may be used in later drug design.

1.4.3 The Vif-induced degradation on APOBEC3G

Because the behaviour of HIV is different in human T cells in the absence or in the presence of Vif, these cells are divided into two groups as ‘permissive cells’ or ‘non-permissive cells’ (Fisher, Ensoli *et al*. 1987, Madani and Kabat 1998, Simon, Gaddis *et al*. 1998). The difference is now widely known as a restrictive capacity to overcome the function of APOBEC3G which has been detailed above. Vif counteracts the anti-viral defence by inducing degradation of some members of APOBEC3 family, including A3DE (Dang, Wang *et al*. 2006), A3F (Wiegand, Doehle *et al*. 2004, Zheng, Irwin *et al*. 2004, Albin, Anderson *et al*. 2013), A3G (Yu, Yu *et al*. 2003, Mehle, Strack *et al*. 2004, Feng, Love *et al*. 2013) and some A3H proteins which originated in west and central Africa (OhAinle, Kerns *et al*. 2008, Tan, Sarkis *et al*. 2009, Zhao, Ishida *et al*. 2012). The mechanism is to neutralize those anti-virus cellular factors by a ubiquitination pathway where an E3 ubiquitin ligase is

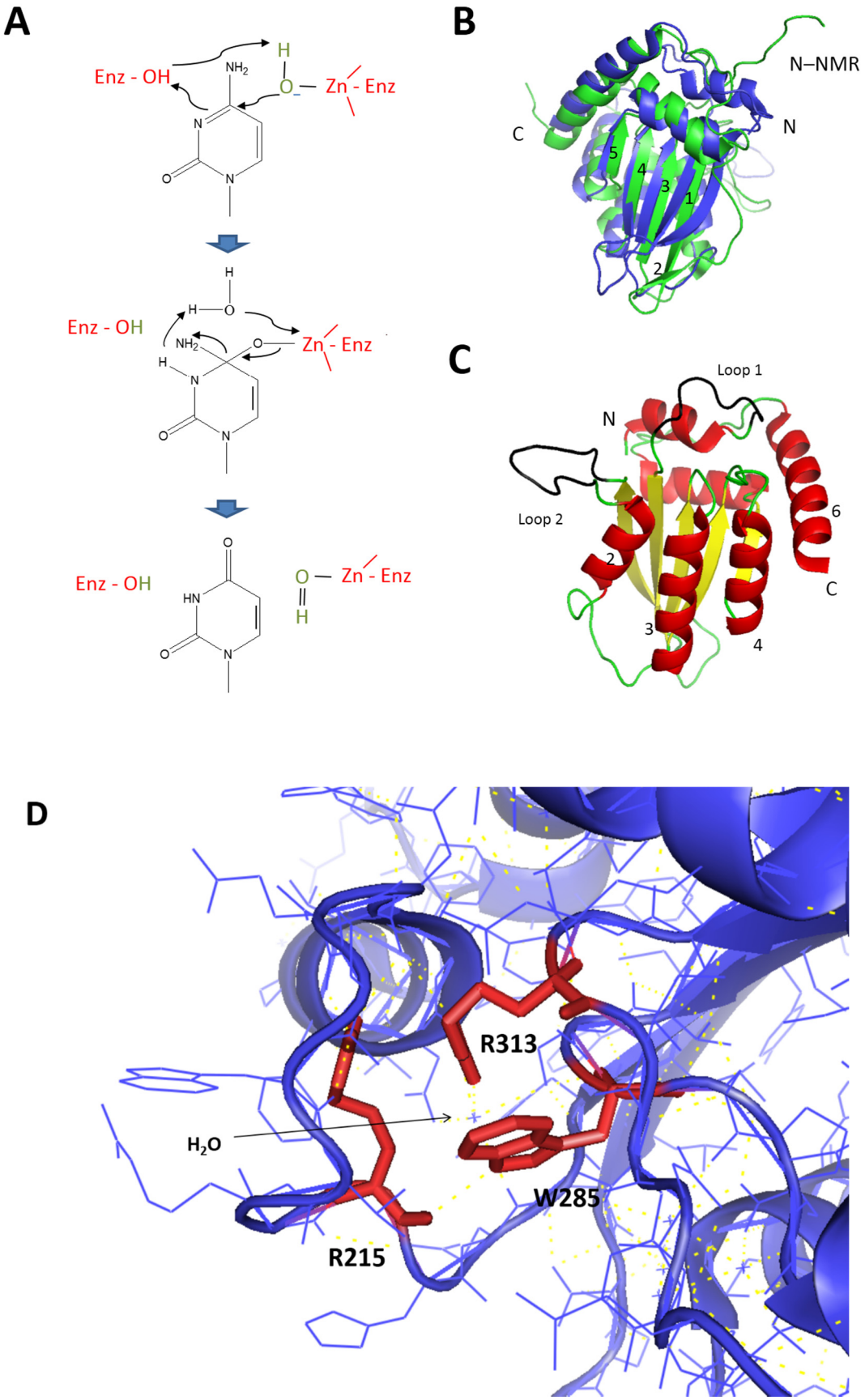

A
Enz - OH
Zn - Enz
B
N–NMR
C
N
Loop 1
Loop 2
C
D
R313
H2O
W285
R215

Figure 1.9: The deamination mechanism and APOBEC3G structures. (A) The proposed mechanism for cytidine deaminases action by APOBEC3G. Enz stands for enzyme. (B) The alignment of the carboxyl terminal solution structure 2JYW (green) and the crystal structure 3E1U (blue). The number of β-strands are presented on the solution structure. The N-terminus and C-terminus are indicated. (C) The crystal structure of C-terminal A3G. The secondary structures are colored differently. The functional domain of loop 1 and loop 2 are highlighted in black. Some α-helix are numbered according to the original paper (Holden, Prochnow *et al*. 2008). (D) Zoom in on the active domain of C-terminal A3G. The side-chains from three critical residues are highlighted in red. The water molecule coloured in blue in the middle are connecting to R313 and W285 via hydrogen bounds. Regenerated from A3G crystal structure 3E1U.

recruited to start the degradation (Conticello, Harris *et al*. 2003, Marin, Rose *et al*. 2003, Sheehy, Gaddis *et al*. 2003, Yu, Yu *et al*. 2003). Co-expression of Vif with A3G improves the degradation rate of the two polyubiquitinated components while Vif is only monoubiquitinated in the absence of A3G (Mehle, Strack *et al*. 2004). However, other studies argue that the correlation between the degradation of APOBEC3G and the rescue of viral infectivity is not direct as expected before and these two functions of Vif are separable (Kao, Goila-Gaur *et al*. 2007).

In the Vif-induced ubiquitination pathway, several specific cellular factors are recruited to form a ligase complex composed of EloB, EloC, Cul5 and Rbx, (Yu, Xiao *et al*. 2004, Luo, Xiao *et al*. 2005, Duda, Borg *et al*. 2008, Mahrour, Redwine *et al*. 2008) in the presence of Vif (figure 1.10A). The core binding factor β (CBFβ), a newly discovered chaperone, is crucial for the folding and stabilizing of Vif (Jager, Kim *et al*. 2012, Zhou, Evans *et al*. 2012, Kim, Kwon *et al*. 2013), the Vif-APOBEC3G interaction, (Zhang, Du *et al*. 2012, Du, Zhao *et al*. 2013) and regulation of host gene expression (Kim, Kwon *et al*. 2013). The formation of the E3 ligase complex results in ubiquitination of the target proteins and neutralization of the anti-viral responses. The Vif sequence and its functional domains are presented in figure 1.10B. In detail, the N-terminus of Vif has been shown to bind to members of the APOBEC3 family (Russell and Pathak 2007, He, Zhang *et al*. 2008). Recent work from John Gross's group shows that the N-terminal residues 1-140 also interact with CBFβ (Kim, Kwon *et al*. 2013). Vif can selectively bind to Cul5 through a zinc-binding motif, $H^{108}x_5Cx_{17\text{-}18}Cx_5H^{139}$ (HCCH motif), a region in the middle of the sequence and upstream of the SOCS binding domain (SOCS-box). Vif without the zinc-binding motif from the non-primate lentivirus has a weaker interaction with the E3 ligase (Luo, Xiao *et al*. 2005, Mehle, Thomas *et al*. 2006).

The SOCS-box of Vif has been known to bind to the EloB-EloC heterodimer along with Cul5. The SOCS-box is C-terminal to the zinc-binding HCCH motif and contains a conserved SLQYLA motif (residues 144-149) called 'BC-box' because of its interaction with EloBC (Yu, Xiao *et al*. 2004, Kobayashi, Takaori-Kondo *et al*. 2005). Mutation of this motif has been demonstrated to result in a functional loss of Vif, which indicates that this domain plays a critical role in Vif (Mehle, Goncalves *et al*. 2004, Mehle, Strack *et al*. 2004, Yu, Xiao *et al*. 2004, Kobayashi, Takaori-Kondo *et al*. 2005). The SOCS-box also has a critical proline-rich

motif (161PPLPS165, PPLPS motif) (Donahue, Vetter *et al*. 2008) downstream of the BC-box, whose biological role is poorly understood regardless of its possible function in forming homomultimers (Yang, Gao *et al*. 2003, Techtmann, Ghirlando *et al*. 2012). The proline-rich motif is found to interact with the EloB C-terminal tail to perform an unknown biological function (Bergeron, Huthoff *et al*. 2010). Mehle's group found that the binding to EloC was regulated by phosphorylation on the Serine-144 site in the SOCS-box domain, implying that this Vif-induced ubiquitination may be relative to other signal regulation pathway (Mehle, Goncalves *et al*. 2004).

Recently there were two reports that a cellular factor named core binding factor β (CBFβ) which usually functions in T-cell differentiation is hijacked by Vif in the ubiquitination pathway, resulting in APOBEC3G degradation and improved viral infectivity (Jager, Kim *et al*. 2012, Zhang, Du *et al*. 2012). It is of note that this factor is able to help Vif fold and become soluble when CBFβ and Vif are co-expressed in *Escherichia coli* (*E. coli*), which significantly highlights that in the structural studies on Vif there is no report on Vif structure except an α-helical BC-box structure (Stanley, Ehrlich *et al*. 2008). Once soluble Vif can be obtained by protein engineering methods, further efforts can use X-ray crystallography or cryo-electron microscopy to solve the structure. The general view of Vif-induced APOBEC3G degradation is shown in figure 1.11.

1.4.4 APOBEC3G-induced HIV-1 evolution and adaption

APOBEC3G suppresses HIV-1 by triggering mutations in the DNA during transcription. In the old days, scientists focused more on the interaction mechanisms of the APOBEC3G-Vif complex because of the potential clinical implications, and ignored the role that APOBEC3G plays in HIV-1 mutation and evolution. A dozen years ago a sequence analysis of HIV-1 from infected people revealed that the subsets of sequences were dominated by G-to-A hypermutation (Janini, Rogers *et al*. 2001), which implies an important biological role of APOBEC3G besides its anti-virus function (Malim and Emerman 2008), i.e. the APOBEC3G deamination capability may facilitate HIV to mutate and evolve, a mechanism by which viruses succeed in evading the host immune system and clinical treatment.

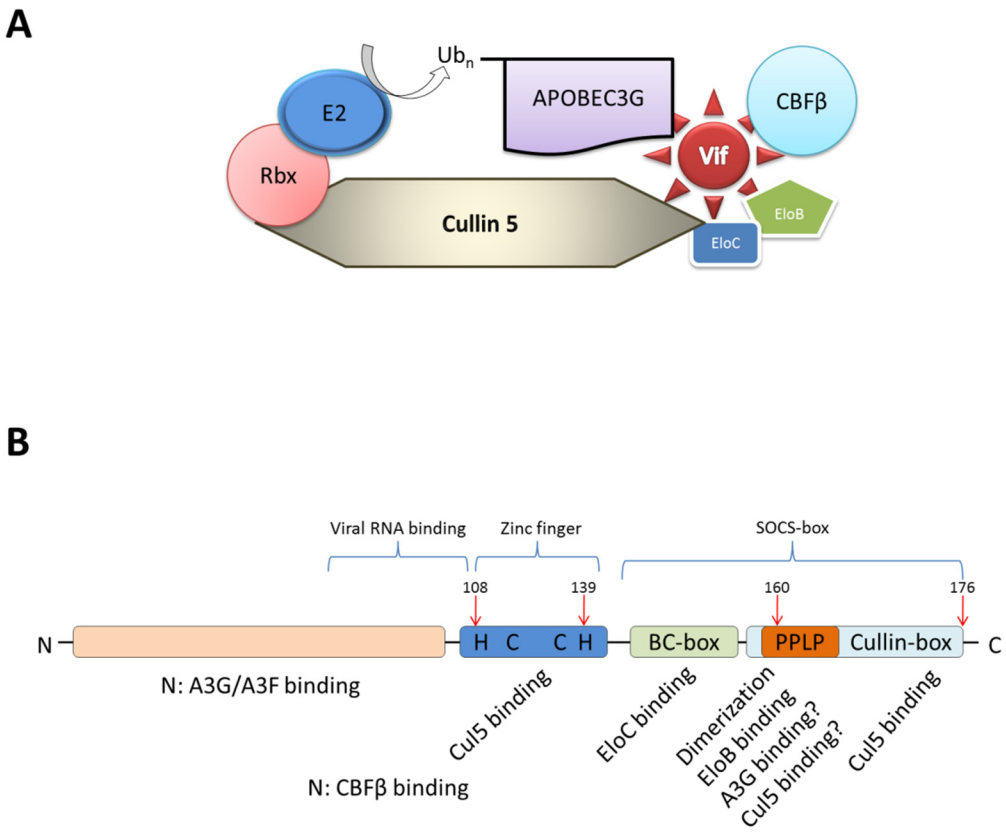


Figure 1.10: The ligase complex and Vif sequence. (A) The Vif-induced ubiquitination complex. The complex composed of Rbx, Cul5, EloB, EloC and Vif is named E3 ligase complex. (B) The Vif sequence (192 amino acids). The functional domains are coloured and indicated respectively.

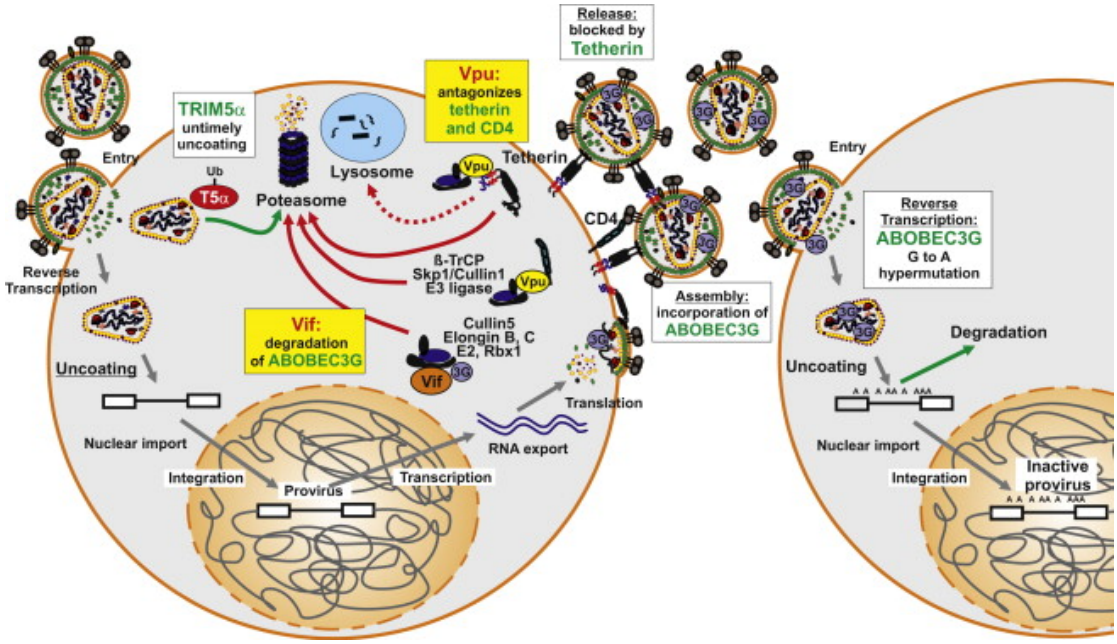


Figure 1.11: The pathway of Vif-induced A3G degradation. Adapted from reference: Immune Evasion and Counteraction of Restriction Factors by HIV-1 and Other Primate Lentiviruses (Kirchhoff 2010).

This hypermutation happens more frequently in a dinucleotide context, especially in the 5′-GG-3′ context which amounts for 95% hypermutations and correlates with higher CD4 counts (Kieffer, Kwon *et al*. 2005, Gandhi, Siliciano *et al*. 2008, Land, Ball *et al*. 2008). This possibly explains that APOBEC3G may help HIV form A-rich genomes (Berkhout and van Hemert 1994), as it has been concluded in the review (Hache, Mansky *et al*. 2006). Although it is still unclear how the APOBEC3 family proteins stimulate pressure on sequence selection during evolution, a study using a newly-developed mathematical model with 3,499 HIV *env* gene sequences from 102 subjects demonstrated that APOBEC3G or APOBEC3F contributed to the sequence diversity (Keele, Giorgi *et al*. 2008).

Importantly, the understanding of the viral evolution and adaption attributed from APOBEC3 family enables us to re-consider the drug resistance that is thought to be related to APOBEC3 proteins. In the clinic, there have been many drug-resistance-related mutations reported in the past decades. One of the resistances to the thymidine analogue zidovudine, a component used for blocking the reverse transcription in a highly active antiretroviral therapy (HAART), was resulted from a G-to-A mutation taking place on an APOBEC3F hotspot, after which an aspartic acid was replaced by an asparagine (Mitsuya, Weinhold *et al*. 1985, Larder and Kemp 1989, Hache, Mansky *et al*. 2006). The effect of Nelfinavir, another typical drug that inhibits HIV protease (Patick, Mo *et al*. 1996), was partially weakened because of amino acid replacement that is obtained by a simple G-to-A DNA mutation as well (Pazhanisamy, Stuver *et al*. 1996, Hache, Mansky *et al*. 2006). Meanwhile, the G-to-A hypermutation virus was reported in 2001, indicating that Vif does not always overcome the function of APOBEC3G, (Janini, Rogers *et al*. 2001). However, Berkhout argued that even in APOBEC3G-negative cells large amounts of G-to-A hypermutation were also detected (Berkhout, Das *et al*. 2001, Berkhout and de Ronde 2004).

Some later studies, including *in vitro* experiments and calculation models, demonstrated that APOBEC3G can facilitate HIV adaption. The HIV-1 provirus mutant with lamivudine-resistance appeared in the presence of APOBEC3G before exposure to a drug. This drug-resistance ability was then acquired by recombination with the mutated provirus (Mulder, Harari *et al*. 2008). Jern, *et al.* developed a computational model to calculate the probabilities of G-to-A mutation targeted by APOBEC3G regarding the ratio of nonsynonymous to

synonymous sequence changes which leaves a footprint during virus generations (Jern, Russell *et al*. 2009). The result revealed that APOBEC3G-targeted HIV-1 genomes experienced mutations at a higher ratio compared with the random control, and that mutants induced by APOBEC3G did acquire the drug-resistance, although it happened at a lower rate, suggesting that APOBEC3G plays a positive role in viral survival and evolution. Another investigation using sequence analysis showed that APOBEC family proteins help HIV-1 escape early immune response by mutating nucleotides on CTL epitope sites (Wood, Bhattacharya *et al*. 2009). The CTL is a component that suppresses HIV in the early immune response.

The discovery that APOBEC3G has a two-sided function has stimulated work producing some encouraging results which are disclosing the secret picture of APOBEC3G-Vif interaction. It is worth mentioning that in Berkhout's case (Berkhout and de Ronde 2004), perhaps it is other cellular factors that lead to the G-to-A hypermutation. On the other hand, there are still lots of questions remaining to be answered. For example, how important is APOBEC3G in viral evolution to help the virus escape immune response and acquire drug-resistance? No doubt APOBEC3G, as a critical anti-virus factor in host cells, can inhibit HIV-1, so what is the balance between the two opposite functions? Besides APOBEC3G and APOBEC3F, do other APOBEC3 family proteins facilitate the evolution of HIV-1? More efforts are required to uncover the complicated mechanisms.

1.5 HIV drug development

In order to suppress HIV-1 levels in plasma, many tools, drugs and treatment strategies have been tried in the past two decades. The most successful treatment is HAART that combines several virus inhibitors working in various steps of viral replication in order to prevent viral mutations that may easily help the viruses escape the effects of one drug.

HAART, however, is not very beneficial due to its high cost and unknown long-term side effects, and most importantly, HAART fails to cure the HIV infection or AIDS because it has no effect on latent integrated provirus. That means these slowly-replicating viruses form stable reservoirs in resting memory $CD4^+$ T cells that will break out sometime later, causing more

severe symptom (Siliciano, Kajdas *et al*. 2003, Richman, Margolis *et al*. 2009, Yang, Xing *et al*. 2009). Therefore, in order to cure the disease, another therapeutic strategy is required to clear provirus in resting T cells and make up for the disadvantages of HAART (Trono, Van Lint *et al*. 2010, Deeks 2012). This strategy now is a leading field in HIV vaccine and drug development.

A report in 2009 from Robert Siliciano's group shed the first light on the discovery of chemical molecules that were able to provoke the 'hiding' provirus (Yang, Xing *et al*. 2009). They established a primary Bcl2-transduced cell strain as a model system and used it to screen chemical-molecule libraries. A small molecule, 5-hydroxynaphthalene-1,4-dione, unlike other activators, is found to be able to reactivate latent viruses in resting $CD4^+$ T cells through ROS and NF-κB pathways without activating the global T cell response. This is a huge advantage because activated T cells will be infected by viruses again which decreases the treatment effects. Recently a group of quinolin-8-ol derivatives have been identified based on initial hits by the same group, which expends the possible drug candidates in the pool (Xing, Bhat *et al*. 2012).

A more encouraging and promising study reported in *Nature* recently demonstrates that this clean-latent-virus strategy can be applied in clinical study and impact HIV-1 latency *in vivo* (Archin, Liberty *et al*. 2012). They chose eight patients who were treated with HAART and in whose body the viruses were suppressed completely. Those patients were given a type of histone deacetylase inhibitor called suberoylanilide hydroxamic acid, known as a vorinostat that has been identified to disrupt HIV-1 latency *in vitro* (Archin, Espeseth *et al*. 2009, Archin, Keedy *et al*. 2009, Contreras, Schweneker *et al*. 2009). In that study, a single dose of vorinostat improved the level of an acetylating biomarker and HIV-1 RNA expression in the resting T cells of each individual, showing that these reactivated cells can be recognized by the immune systems. This proof-of-concept study indicates that stimulated HIV-1 reservoirs can be therapeutically targeted and then removed by novel drugs and feasible strategies in the future. Besides the drug development, recently the development of broadly neutralizing antibodies has also achieved a huge success (Wu, Yang *et al*. 2010, Wu, Zhou *et al*. 2011), which has been highlighted in several latest reviews (Haynes and McElrath 2013, Mascola and Haynes 2013, Stephenson and Barouch 2013).

1.6 Aim of this thesis

Due to problems with over-expressing, purifying and crystallizing the full-length folded soluble Vif and APOBEC3G (Barraud, Paillart *et al*. 2008, Engelman and Cherepanov 2012), it is still unclear from a structural biological viewpoint how Vif recruits cellular factors and interacts with APOBEC3G to stimulate degradation. Most reports until now have been on the properties of Vif despite only a structural study on the α-helical BC-box domain (Stanley, Ehrlich *et al*. 2008). Therefore, the purpose of this project is to elucidate Vif structure and function by biophysical methods including NMR, crystallography, mass spectrometry (MS), isothermal titration calorimetry (ITC), and cellular biological assays.

The Vif SOCS-EloBC structure is reported first. Following our previous research (Bergeron, Huthoff *et al*. 2010), more NMR data allows us to solve the 36 kDa complex structure, revealing the interface between the proline-rich motif of Vif and EloB carboxyl terminus. Moreover, biophysical and biological assays characterize the role of residues on the proline-rich motif. The results suggest that it is a weak interaction between the proline-rich motif and EloB subunit involving van der Waals forces.

In the second part, we are more interested in purifying and crystallizing the full-length Vif after CBFβ was reported to function in the Vif-induced degradation pathway (Jager, Kim *et al*. 2012, Zhang, Du *et al*. 2012). The complete protocol of protein production is presented, and solubility tests through different methods are discussed to explain the conditions used for purification. Interestingly, the behaviour of two CBFβ isoforms are different during the sample preparation, indicating that the carboxyl terminus of the shorter isoform participates in helping Vif fold in cells. In the end, crystallization results are described and further suggestions are given.

CHAPTER 2: MATERIAL AND METHODS

2.1 Molecular cloning

The plasmids for Vif SOCS-box peptide and EloBC were made by Dr. Julien Bergeron (Bergeron, Huthoff *et al*. 2010). CBFβ isoform I and II cDNAs were synthesized by Genscript and BlueHeron respectively. Vector pRSFDuet was purchased from Novagen.

Polymerase chain reactions (PCR) were performed using the Phusion® DNA polymerase kit (Finnzyme) in a 50 µl containing 1 unit of polymerase, 10 mM dNTPs (each), 10-50 ng DNA template and 10-50 pmol of 5′ or 3′ primer in 'High Fidelity' reaction buffer. Sequences of primers used for different constructs are listed in appendix I. A general program used for PCR is described in table 2.1.

Table 2.1, PCR program used for molecular cloning

Step	Temperature	Time
1	95 ℃	2 minutes
2	95 ℃	30 seconds
3	55 ℃	30 seconds
4	68 ℃	1 minute
5	Repeat step 2-4	30 cycles
6	68 ℃	10 minutes
7	4 ℃	hold

Plasmids, PCR products and vectors were digested by restriction endonucleases from New England Biolabs (NEB). Reactions were carried out in a 40 µl volume containing 4 µl enzyme buffer depending on the endonucleases, 4 µl 10% bovine serum albumin (BSA), 1000 ng DNA, and 1 µl of each required enzymes. The reaction was incubated at 37 ℃ for 3 hours or overnight.

Plasmids and PCR products or restriction digests were analyzed and separated by mixing 5 µl of sample with 1 µl of 6* loading buffer (Abgene) on a 1-1.2% agarose gel containing Tris/Borate/EDTA buffer (Fisher) and 0.5 µg/ml ethidium bromide. Ladders from NEB were used to estimate the size of the DNA of interest. Gels were run at 100 V, and visualized under the ultraviolet trans-illuminator (Biorad).

The products of PCR reactions or digested DNA separated on agarose gel were purified via a QIAquick® purification kit (QIAGEN), according to the manufacturer's instructions. The DNA was re-suspended in around 50 µl H_2O.

In order to insert DNA fragments into expression vectors, 2 µl digested vector and 6-10 times the amount of DNA fragments were mixed in a total reaction volume of 20 µl containing T4 ligase buffer and 1 µl T4 DNA ligase (NEB). The transforming reaction mix was incubated at 4 ℃ overnight. 10 µl of ligation reaction was transformed into 80 µl volume of XL-blue competent cells. The solution mix was incubated on ice for 45 min, followed by a heat shock at 42 ℃ for 1 min. After recovering in 250 µl Luria-Bertani broth (LB) at 37 ℃ for 45 min, transformed cells were plated on LB-agarose plates containing appropriate antibiotics. After 18 hours, colonies were picked and grown at 37 ℃ for 12 hours in 5 ml LB plus antibiotics. Plasmids were extracted using the QIAprep® spin miniprep kit (QIAGEN). Final constructs were confirmed by PCR reactions or by digest tests, as well as by sequencing (MWG-biotech).

2.2 Bioinformatic analysis

Nucleotide blast was performed on the BLAST server (blast.ncbi.nlm.nih.gov/Blast.cgi) on the National Center for Biotechnology Information website to validate the cDNA sequences from molecular cloning. Protein sequence alignment and relative conservation analysis were performed using Clustal-W with default settings (Larkin, Blackshields *et al*. 2007).

Secondary structures were predicted by PsiPred (Jones 1999). The prediction of flexibility was carried out using Random Coil Index (RCI) from the backbone assignments (Cα CO, Cβ, N, Hα, NH) and employing an RCI web server (Berjanskii and Wishart 2005). Torsion angles were predicted by TALOS+ from backbone assignments (Shen, Delaglio *et al*. 2009, Wassenaar,

van Dijk *et al*. 2012).

The characteristics of residual sequences, including pI, positive/negative charges and Scan Mass Spectrometry results were acquired from Protein Calculator provided by Scripps Institute (www.scripps.edu/~cdputnam/protcalc.html). Hydrophobic statistics, molecular weight and amino acid abundance of sequences were analyzed by a self-written script that is attached and explained in appendix II. Extinction coefficient values, grand average of hydropathicity and instability index (Guruprasad, Reddy *et al*. 1990, Walker 2005) were obtained from ProtParam, Expasy (http://web.expasy.org/protparam/).

2.3 Protein expression and purification

2.3.1 Protein expression

Vif SOCS-box and EloBC were expressed as described previously (Bergeron, Huthoff *et al*. 2010). For the Vif/EloBC/CBFβ tetramer, plasmids pETDuet with EloBC constructs and pRSFDuet with Vif/CBFβ constructs were co-transformed into BL21 (DE3) pLysS competent cells, followed by plating cells on LB-agarose dishes containing 50 ng/ml Kanamycin and 100 ng/ml ampicillin and grown at 37 ℃ overnight. One colony was picked and grown in 3 ml LB with appropriate antibiotics at 37 ℃ for 12 hours. One volume of bacteria was mixed with one-volume of sterilized 50%-glycerol LB media to produce a glycerol stock which was stored at -80 ℃. All bacteria used for protein expression and purification later were from the glycerol stocks.

Cells from -80 ℃ were grown in 3 ml LB at 37 ℃ for 12 hours. 10 μl of recovered bacteria was included to fresh 5 ml LB containing 0.1 mM $ZnCl_2$ and grown at 37 ℃ for another 12 hours. Next day the overnight culture was added into 800 ml LB containing 0.1 mM $ZnCl_2$ and appropriate antibiotics. Expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM when OD_{600} reached 0.8, the temperature was decreased to room temperature. Proteins were then expressed at 16 ℃ for an additional 20 hours. Cells were collected by 6,000 rpm centrifugation at 4 ℃ for 20 min.

Pellets were stored at -20 ℃. Cell lysis after induction was not usually taken for SDS-PAGE gel analysis because the amount of protein in each cell was very low due to the induction temperature and IPTG concentration. All the media described above contained 50 ng/ml kanamycin and 100 ng/ml ampicillin.

2.3.2 Protein solubility test

1 ml of cells recovered from glycerol stock were added to 100 ml LB media containing 0.1 mM $ZnCl_2$, 50 ng/ml kanamycin and 100 ng/ml ampicillin. Proteins were induced for 4 hours with 0.5 mM IPTG at 37 ℃ or low temperature as described above. After induction the media was distributed into 10 aliquots of 10 ml each. Bacteria pellets were spun down at 4,500 rpm for 20 min and re-suspended in 500 µl of different buffers with pH and salt gradient. Samples were sonicated at 30% amplitude using a Sonifier®250 (Branson) for 2 min with a pulse program of 5-sec on/10-sec off. Cell lysis was centrifuged at 17,000 rpm for 30 min. 20 µl of cell lysis and supernatant were added to 2* gel loading buffer. 10 µl of each sample was run on a 15% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) or Tricine gel, displayed with Coomassie blue. 1 µl of each sample was used for western blot with Vif antibody.

2.3.3 Nickel-nitrilotriacetic acid affinity purification

The purification of Vif SOCS-box and EloBC has been described before (Bergeron, Huthoff *et al.* 2010). Pellets for the Vif/EloBC/CBFβ tetramer were resuspended in Vif re-suspending buffer (20 mM phosphate pH 7.0, 100 mM NaCl, 20 mM imidazole, 10% glycerol, 8 mM 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate hydrate (CHAPS), Roche protease inhibitor cocktail, 0.05% sodium azide) and sonicated by Sonifier®250 for 10 min with a pulse program of 5-sec on/10-sec off. The cell lysis was centrifuged at 17,000 rpm at 4 ℃for 1 hour. The supernatant was loaded onto the equilibrated Ni-NTA agarose beads (QUIAGEN) and the flow-through was then re-loaded onto the column, which was subsequently washed with approximately 200 ml of washing buffer (20 mM phosphate pH 7.0,

250 mM NaCl, 40 mM imidazole, 0.05% sodium azide). Proteins were eluted into elution buffer (20 mM phosphate pH 7.0, 250 mM NaCl, 200 mM imidazole, 8 mM CHAPS, Roche protease inhibitor cocktail, 0.05% sodium azide) and stored at 4 ℃. The entire process of purification was conducted in the cold room.

2.3.4 SDS-PAGE analysis

15% SDS-PAGE was used to analyze protein samples. Running gels (375 mM Tris-HCl pH 8.8, 0.1% ammonium persulphate, 0.1% SDS, 0.004% N,N,N′,N′-Tetramethylethylenediamine (TEMED), 15% acrylamide) was cast in ATTO (Japan) system and covered with butanol until the gel had polymerized. The butanol was discarded after 30 min and the spacer gel (125 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% ammonium persulphate, 0.004% TEMED, 4% acrylamide) was loaded on the running gel. Between 3 µl and 20 µl of the samples were loaded onto the gel depending on samples, and run at 120 V for 2 hours in running buffer (27.6 mM Tris, 0.2 M glycine, 0.1% SDS).

2.3.5 Tricine gel analysis

Tricine is an organic compound made from tris and glycine where the word 'Tricine' derives from. The Tricine gel system is a modification of traditional SDS-PAGE for low molecular weight proteins. In solution Tricine is more negatively charged than glycine, making the small-molecule highly charged and thereby move faster in the gel. This allows small proteins whose molecular weight (MW) is between 1 kDa to 100 kDa to be separated (Schagger and von Jagow 1987, Lesse, Campagnari *et al*. 1990).

Tricine gel was used to analyze the tetramer sample owing to the slight difference in protein MW between Vif and CBFβ. The running gel contained 15% acrylamide, 1.0 M Tris-HCl pH 8.45, 0.1% SDS, 10% glycerol, 0.033% ammonium persulphate and 0.004% TEMED. The spacer gel contained 4% acrylamide, 0.6 M Tris-HCl pH 8.45, 0.06% SDS, 0.033% ammonium persulphate and 0.004% TEMED. Running buffer covering the top of the gel contained 0.1 M Tris, 0.1 M Tricine pH 8.25, and 0.1% SDS. Running buffer in the bath was made from 0.2 M

Tris pH 8.9. The gel was run at 100 V for 5 hours and stained overnight.

2.3.6 Protein preparation for Nuclear Magnetic Resonance Spectroscopy

Bacteria containing the Vif SOCS-box construct and the EloBC construct were recovered in 5 ml LB media overnight. Cells were collected by 4,000 rpm centrifugation for 10 min and the media was discarded. Pellets were re-suspended in 100 ml M9 minimal media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 1 g/L $^{15}NH_4Cl$, 4 g glucose) containing ^{15}N or ^{13}C according to experiment requirements and grown overnight. The small culture was transferred to 900 ml minimal media containing ^{15}N or ^{13}C on the next day. Proteins were induced by IPTG to a final concentration of 1.0 mM at OD_{600} 0.8 for four hours. Cells were collected as described above. For deuterated samples, recovered bacteria were grown in 5 ml minimal media culture overnight. The culture was then grown in a small volume of 30% and 70% D_2O minimal media before 1L of 100% deuterated minimal media was used to express proteins.

1-^{13}C glutamine and 1-^{13}C proline-labeled proteins were obtained in 100% deuterated M9 minimal media containing $^{15}NH_4Cl$ and non-labeled glucose. 0.2 g of each labeled amino acid were re-suspended in 10 ml minimal media and added to the cell culture just before induction.

All labeled proteins were purified as described above.

2.4 Biophysical assays

2.4.1 Analytical gel filtration study

Size exclusion column purification of SOCS-box and EloBC have been detailed in Dr. Julien Bergeron's paper (Bergeron, Huthoff *et al*. 2010). As for the SOCS-EloBC complex, 1 volume of labeled component was mixed with 1.2 volume of unlabeled component and concentrated to 1 ml. Sizing exclusion column GE Superdex 75 10/300 GL was equilibrated with NMR buffer (10% D_2O, 50 mM sodium-phosphate buffer pH 6.7, 0.05% sodium azide)

before using. A final sample of the 1:1 complex sample was collected usually from fraction 21 to 25. Protein concentration was measured by UV at 280 nm wavelength on NanoDrop VD-1000 spectrophotometer.

As for the Vif-EloBC-CBFβ tetramer, the whole gel filtration system was sitting in a cold cabinet at 4 ℃. Size exclusion column Superdex 200 10/300 GL was equilibrated with 30 ml Vif final buffer (10 mM phosphate buffer pH 7.0, 250 mM sodium chloride, 0.02% sodium azide, 8 mM CHAPS) before the experiment. Protein samples from a Ni-NTA column were concentrated to 2 ml. In each run 1 ml samples was loaded onto the column. The running speed was set at 0.5 ml/min per fraction.

In order to calibrate the system and to obtain a guide on the behaviour of various molecules with different MW, a group of standard components were diluted in 1 ml of the final buffer and tested on the Superdex 200 gel filtration column at 4 ℃. Each component and its MW is listed in table 2.2. Target samples of the tetramer usually came out at approximately 14-ml washing volume, or around fraction 24.

Table 2.2, Protein ladder for gel filtration analysis

Name	Molecular weight / Da	Peak position / ml
Thyroglobulin, bovine	669,000	7.6
Apoferritin, horse spleen	443,000	11.3
β-Amylase, sweet potato	200,000	12.4
Albumin, bovine serum	67,000	14.4
Carbonic Anhydrase, bovine erythrocytes	29,000	16.5
lysosome	11,000	19.1
Aprotinin, Bovine lung	6,500	22.2

2.4.2 Dynamic light scattering

Protein samples and standard control BSA were diluted to approximately 0.2 mg/ml (by

BioRad) in the final buffer. The sample room of the machine was cooled down to 10 ℃ for 15 min before use. A 12 µl sample was added into 12-µl Quartz cuvette after a hard spin at 13,000 rpm at 4 ℃ for 10 min. The molecular diameter was detected and converted to molecular weight by the Dynapro 99 manufactured by **PROTEIN**SOLUTION.

2.4.3 Differential condition scanning

Differential condition scanning is also called thermal shift assay (TSA). 13 µl of test buffer was placed into 96-well PCR plates, followed by adding 1.5 µl protein samples (1 mg/ml by BioRad) into each condition. The original Dye SYPRO® Orange protein gel stain from Life Technologies™ was diluted at 1:31 ratio before loading 0.5 µl diluted dye into conditions. All processes were performed on ice. The sealed PCR plate was put in Stratagene MX3005P qPCR machine and incubated at 25 ℃ for 10 min, followed by a gradient temperature rising to 95 ℃ at 1 ℃/min.

2.4.4 Analytical ultracentrifugation

The Vif tetramer was concentrated to approximately 1 mg/ml measured by NanoDrop and diluted into 0.7 mg/ml and 0.5 mg/ml to obtain three concentrations. Each sample was run in a Beckman Coulter XL-A centrifuge. The experiments and analysis were performed by Dr. Rebecca Beavil at the Randall Division.

2.4.5 Isothermal titration calorimetry

Protein samples are purified from a sizing column and dialysed into a binding buffer, and concentrated to the required concentrations. 300 µl of 0.02 mM SOCS-box or mutants were placed in the chamber of MicroCal® ITC200 calorimeter and 2 µl aliquots of 0.2 mM EloBC was injected at 4 min intervals. The data were plotted using a single binding site model by the ITC software ITC200 with the Origin software provided with the calorimeter. The concentration of SOCS-box was measured by NanoDrop spectrophotometer and that of EloBC was measured

by the BioRad® PROTEIN ASSAY Dye Reagent Concentrate.

2.4.6 Mass spectrometry

In order to prepare the sample for Time-of-Flight mass spectrometry (ToF-MS), purified protein was mixed with loading buffer and run on a large SDS-PAGE. Bands stained by Coomassie blue ware cut and kept in water. The experiments and analysis were performed by Dr. Steve Lynham at the Proteomics Facility, Kings College London, Denmark Hill Campus.

2.5 Nuclear magnetic resonance (NMR) theory

2.5.1 NMR mathematical and physical principles

NMR was first described by the American-Hungarian physicist Isidor Rabi who was awarded the Nobel prize for the discovery (Rabi, Millman *et al*. 1939) in molecular beam experiments. The overall spin of a nucleus is defined by a scalar quantity called the spin quantum number. All isotopes containing an odd number of nuclides have a non-zero spin whereas an isotope with an even number of nuclides has spin 0. Isotopes which are widely used for protein NMR are listed in table 2.3 below.

A non-zero spin is defined by a non-zero magnetic moment μ according to equation (1):

$$\mu = \gamma S \tag{2.1}$$

where γ is the gyromagnetic ratio and S is the angular momentum. This means each moment is quantized. It is the magnetic moment that determines whether a spin system can be observed by NMR. In an external magnetic field, a spin system usually has two states, either parallel or anti-parallel with respect to the field. If the z axis is chosen for B_0, the energy E is given by equation (2):

$$E = -\gamma \hbar m B_0 \tag{2.2}$$

Obviously, as for m=1/2 illustrated in figure 2.1, it is easy to calculate that the energy difference

between two states, namely

$$E = \gamma \hbar B_0 \tag{2.3}$$

The populations of parallel and anti-parallel spins have a Boltzmann distribution, in which case the population rate of up-state β spin N_β to the down-state α spin N_α is:

$$N_\alpha / N_\beta = e^{-\Delta E / kT} \tag{2.4}$$

where T is the absolute temperature in Kelvin and k is a constant whose unit is J/K.

Table 2.3, Spin values for various nuclei

Isotopes	Spin
^{1}H	1/2
^{2}D	1
^{13}C	1/2
^{15}N	1/2
^{17}O	1/2
^{31}P	1/2
^{33}S	3/2

In NMR, relaxation is an important term that describes the process of a spin returning the equilibrium state from a non-equilibrium state. There are two main relaxation processes characterized by the relaxation time constant T_1 and the relaxation time constant T_2. T_1, the spin-lattice relaxation time, is the decay constant with which the excited z component vector, M_z, reverts backs to its thermal equilibrium state M_0. It is defined by the equation (5):

$$M_z(t) = M_0 + [M_z(0) - M_0] \cdot e^{-t/T_1} \tag{2.5}$$

T_2 is called the spin-spin relaxation time, or transverse relaxation time and corresponds to a dephasing of magnetization between spins. This decay constant is defined by the equation (6):

$$M_{xy}(t) = M_{xy}(0) \cdot e^{-t/T_2} \tag{2.6}$$

In most situations, T_1 is larger than T_2, but not in principle (Levitt 2008).

In 1948, Nicolaas Bloembergen, *et al.* presented a theory which accurately explained many experiments, taking into account the effect of the tumbling motion of a large molecular (Bloembergen, Purcell *et al*. 1948). In solution, molecules tumble at a certain rate determined by temperature, viscosity of solution and molecule weight. This motion is termed the correlation function G:

$$G(t) = G(0) \cdot e^{-t/\tau_c} \tag{2.7}$$

Both T_1 and T_2 decay constants are related to the overall correlation time τ_c (figure 2.2).

2.5.2 NMR spectroscopy

In NMR, active nuclei will have an absorption at a frequency characteristic of each isotope which is termed the 'resonance frequency'. The energy of absorption and the signal acquired later are in proportion to the magnitude of external magnetic field, which means that higher field NMR spectrometers, such as 700 megaHz (MHz), 800 MHz or even 1 GHz are much more suitable for large biological macromolecules.

Signals are acquired from NMR detectors in the form of a radiofrequency pulse with ranges of frequency centred on the Lamor frequency, followed by a signal decay (known as free induction decay, FID). This time-domain signal must be converted to a frequency-domain formula by Fourier transform:

$$f(v) = \int g(t) e^{-2\pi i t v} dt \tag{2.8}$$

where *v* represents a particular frequency and t is the time of FID.

The chemical shift δ is a measured unit which is related to the resonance frequency, used for describing the dependence of energy levels in a molecule (Kemp 1987, Balci 2005). Different types of isotopes, e.g. ^{1}H, ^{13}C and ^{15}N that are widely used in protein research, usually behave differently depending on many factors, such as coupling, binding partner, proton distance and of course, the local magnetic field which means in practice every single atom has its own unique chemical shift from which we can derive the structural information.

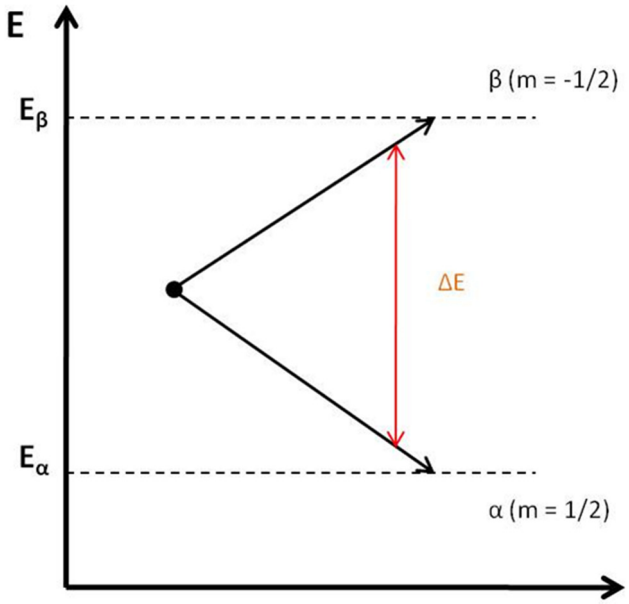


Figure 2.1: The energy difference with m = ±1/2. This difference will change local magnetic flux density B_0.

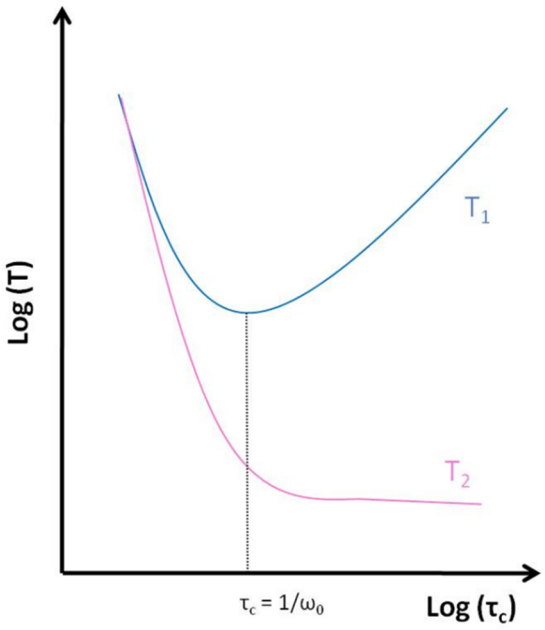


Figure 2.2: The relationship between correlation time (τ) and two relaxation time T1 and T2. ω_0 is the Larmor frequency.

The chemical shift δ is defined as:

$$\delta = \frac{\omega - \omega_0}{\omega_0} \cdot 10^6 \tag{2.9}$$

where ω is called Larmor frequency of a spin and ω_0 is the resonance frequency of a reference substance, usually tetramethylsaline (TMS) (Keeler 2005).

Some important information is acquired from the scalar coupling or J-coupling between NMR active nuclei rising from the interaction between different spin states, which leads to a splitting of the NMR signals (Keeler 2005). For instance, for 1/2 spin systems, coupling to an equal nuclei will be split into n+1 multiple signals whose intensity follows Pascal's triangle (table 2.4):

Table 2.4, Pascal's triangle for a J-coupling system

Multiplicity	Relative intensity
Singlet	1
Doublet	1:1
Triplet	1:2:1
Quartet	1:3:3:1
Quintet	1:4:6:4:1
|	|

This splitting pattern is dependent on the nature of atoms as well as the connecting bond. Furthermore, magnetization on a spin, typically ^{1}H, can be transferred to its neighbouring spin, usually ^{13}C and ^{15}N, via their covalent bonds by the use of precisely-designed pulse sequences of radio-frequency pulses. This is called correlation spectroscopy (COSY), a kind of two-dimensional (2D) NMR spectroscopy. The intensities of chemical shifts in a 2D spectrum can be represented by a third dimension (3D).

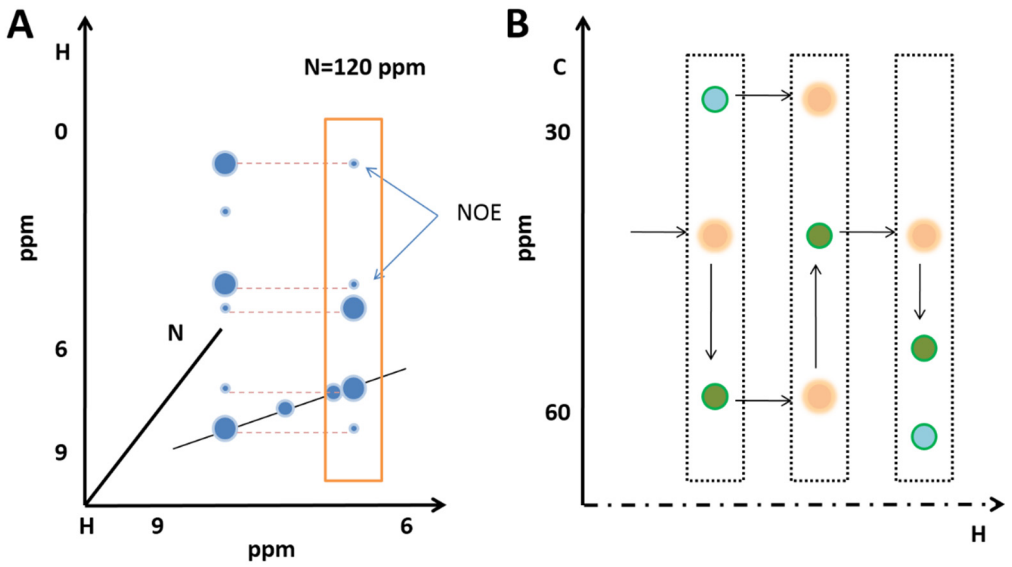


Figure 2.3: 3D NMR spectra cartoons. (A) In a particular plane, some peaks (small circles) from neighbouring spin system can be seen in the spin strip. The intensity of the peaks can be used to calculate the distance between the two spins. (B) In a HNCACB spectrum, Cα and Cβ peaks from current (blue and green) and previous residues that usually give peaks with low intensity (yellow) are visible in one strip, which enables one to assign all peaks in order. The intensity of peaks corresponding to current residues are usually stronger because of the energy transfer distance. Some unique amino acids will have special carbon ppm, e.g. Gly (middle) and Ser/Thr (right).

There are several merits to 2D or 3D NMR. First, more dimensions help to better separate peaks for a many-spin system, especially for high molecular-weight molecules, therefore peaks will not overlap each other, which means it is much easier to conduct assignment. Besides, heteronuclear NMR allows one to analyse the interaction among different types of isotopes. Finally, multi-dimensional NMR, specifically, the nuclear Overhauser effect (NOE) spectroscopy (NOESY), can provide more information on side-chains and distance restraints for structural studies (figure 2.3A).

2.5.3 Fundamentals of spectral assignment

NMR signals are detected and acquired by special probes in the spectrometers. As described briefly above, peaks in 'chemical shift' units are converted from resonance frequency and are displayed on the control interface. For a particular protein structure, each isotope in the protein, such as ^{1}H, ^{13}C and ^{15}N, has its own resonance frequency determined by the intrinsic non-zero spin and the other spin systems around it, which means a peak with unique ^{1}H, ^{13}C and ^{15}N chemical shift on a spectrum always comes from the same residue. Based on this idea, a general backbone assignment protocol is used to start identifying amino acid residues to which peaks are corresponding.

^{1}H-^{15}N heteronuclear single quantum coherence spectrometry (HSQC) is usually regarded as the starting spectrum or standard spectrum (Cavanagh 2007). 3D ^{13}C spectra, including HNCACO, HNCACB and CBCACONH are used for the backbone assignment (figure 2.3B) because sequential backbone carbons, i.e. Cα and Cβ, can be identified by following the carbon chemical shifts in a strip where peaks from previous and current amino acids appear. CC(CO)NH (Bax and Ikura 1991, Grzesiek and Bax 1993), total correlation spectrometry (TOCSY) (Bax, Clore *et al*. 1990) and NOESY (Marion, Driscoll *et al*. 1989) are usually applied for side-chain assignments.

In this project, the side-chain assignments for the Vif SOCS-box monomer was based on NOESY and TOCSY spectra together with Dr. Julien Bergeron's backbone assignment (unpublished). HNCO, HCCH-COSY, HCCH-TOCSY, ^{15}N-edited NOESY-HSQC, ^{13}C-edited NOESY-HSQC spectra were recorded in order to assign SOCS-box sequence in complex.

HNCO, HNCACO, CBCACONH, HNCA, HNCACB, HCCH-TOCSY, ^{1}H-^{15}N TROSY-HSQC and ^{13}C-edited NOESY-HSQC spectra were taken into account for EloBC assignments. Jeffrey Babon's published EloBC assignments (Babon, Sabo *et al*. 2008) were referred to for comparison.

2.6 NMR spectroscopy experiments

2.6.1 NMR acquisition and processing

All NMR samples were prepared as stated above. Sigma NMR tubes with the required 550 µl sample volume were used for each experiment. All the experimental details are listed in table 2.5. TOPSPIN_3.0 was used for acquisition on a Bruker machine. Software VNMR was used for acquisition on a Varian machine. Raw data from NMR spectrometers was processed and transformed by NMRPipe (Delaglio, Grzesiek *et al*. 1995). Assignment, NOESY distance restraint calculations, dihedral constraint calculation and paramagnetic relaxation enhancement measurements were executed by using the CcpNmr suite (Fogh, Boucher *et al*. 2005, Vranken, Boucher *et al*. 2005).

2.6.2 NMR titration

^{15}N-labelled SOCS-box NMR sample was prepared as usual with a final concentration of 0.2 mM in 500 µl phosphate buffer. Unlabelled EloBC sample was concentrated to 0.5 mM and added to SOCS-box sample with a gradient ratio (EloBC : SOCS) 0.25:1, 0.5:1, 0.75:1, 1:1 and 1.2:1. The pulse-program was copied to each experiment to keep all parameters the same. As for the perturbation studies, each EloBC mutant sample was prepared in the same way as wild type and divided into two aliquots. Purified SOCS-box peptide was added to one aliquot at a 1.2:1 ratio and the other was topped up to the same volume as aliquot one with NMR buffer.

Table 2.5 Summary of NMR spectroscopy experiments (Up to April 30, 2013)

Date	Location/Lab	Protein Name	[conc] / mM	T / ℃	Labelling	Spectrum	Spectrometer
30/10/2010	NMRC[a], Mill Hill	SOCS -EloBC[b]	0.16	25	^{15}N SOCS	3D-NOESY	Bruker 700 MHz
27/11/2010	NMRC, Mill Hill	SOCS -EloBC	0.2	25	^{15}N both	3D-NOESY	Bruker 700 MHz
13/12/2010	NMRC, Mill Hill	SOCS -EloBC	0.2	25	^{15}N SOCS	3D-NOESY	Varian 800 MHz
17/01/2011	Wolfson Wing, KCL	SOCS -EloBC	0.4	25	^{13}C, ^{15}N SOCS	COSY, TOCSY, C/H-CCONH[c]	Bruker 700 MHz
20/01/2011	NMRC, Mill Hill	SOCS and EloBC	0.16	4[d]	^{15}N SOCS	Titration, T gradient, 3D-NOESY	Bruker 700 MHz
01/02/2011	Wolfson Wing, KCL	SOCS -EloBC	0.4	-	^{13}C, ^{15}N SOCS[e]	T gradient, 25, 30, 37	Bruker 700 MHz
10/02/2011	NMRC, Mill Hill	SOCS monomer	0.5	25	^{13}C, ^{15}N SOCS	^{13}C-HSQC, ^{15}N-TOCSY, NOESY	Varian 800 MHz
08/03/2011	Wolfson Wing, KCL	SOCS -EloBC	0.2	30	^{15}N EloBC	HNHA	Bruker 700 MHz
31/03/2011	NMRC, Mill Hill	SOCS -EloBC	0.2	25	^{13}C, ^{15}N SOCS	^{13}C-NOESY, HNCO	Bruker 700 MHz
05/05/2011	Wolfson Wing, KCL	SOCS -EloBC	0.4	30	^{15}N both	HSQC in 20/50 mM buffer	Bruker 700 MHz
22/05/2011	Wolfson Wing, KCL	SOCS -EloBC	0.25	30	^{13}C, ^{15}N EloBC	^{13}C-TOCSY, NOESY	Bruker 700 MHz
14/06/2011	Wolfson Wing, KCL	SOCS -EloBC	0.27	37	^{15}N SOCS	`^{15}N-HSQC for RDC	Bruker 700 MHz
23/06/2011	NMRC, Mill Hill	SOCS -EloBC	0.1	37	^{15}N SOCS	^{15}N-HSQC for RDC	Varian 800 MHz
		SOCS -EloBC	0.4	25	^{15}N both	^{15}N-NOESY	

21/07/2011	NMRC, Mill Hill	SOCS -EloBC	0.4	25	^{15}N both	^{15}N-TOCSY, 150 ms NOESY	Bruker 700 MHz
19/01/2012	NMRC, Mill Hill	SOCS -EloBC	0.37	30	L, Q on EloBC[f]	Trosy-^{13}C-HNCO	Varian 800 MHz
14/02/2012	NMRC, Mill Hill	SOCS -EloBC	0.2	30	L, Q on EloBC	Trosy-^{13}C-HNCO	Bruker 700 MHz
15/02/2012	Wolfson Wing, KCL	SOCS -EloBC	0.2	25	^{15}N EloBC[g]	^{15}N-HSQC for RDC	Bruker 700 MHz
10/05/2012	NMRC, Mill Hill	MIII-EloBC[h]	0.1	30	^{15}N SOCS	^{15}N-TOCSY, NOESY	Varian 800 MHz
11/05/2012	Wolfson Wing, KCL	SOCS MIII[i]	0.21	30	^{15}N SOCS	^{15}N-TOCSY, HNHA	Bruker 700 MHz
19/07/2012	NMRC, Mill Hill	HIV2 SOCS	0.02	25	^{13}C, ^{15}N	HNCA	Bruker 700 MHz
30/08/2012	NMRC, Mill Hill	HIV2 SOCS	0.04	25	^{13}C, ^{15}N	HNCA	Varian 800 MHz
17/10/2012	Wolfson Wing, KCL	SOCS *-EloBC	0.055	25	^{15}N EloBC	^{15}N-HSQC for paramagnetic	Bruker 700 MHz
30/10/2012	NMRC, Mill Hill	HIV2 SOCS	0.1	25	^{15}N	^{15}N-NOESY	Bruker 700 MHz
08/11/2012	NMRC, Mill Hill	HIV2 SOCS	0.05	25	^{13}C, ^{15}N	HNCOCA	Bruker 700 MHz
20/12/2012	NMRC, Mill Hill	SOCS	0.2	25	^{13}C, ^{15}N	^{13}C-TOCSY[j]	Varian 800 MHz
15/01/2013	NMRC, Mill Hill	SOCS	0.2	25	^{13}C, ^{15}N	HCCCONH[k]	Bruker 700 MHz
17/01/2013	Wolfson Wing, KCL	SOCS -EloBCM[l]	0.05	25	^{15}N EloBC	^{15}N-HSQC for binding test	Bruker 700 MHz
25/01/2013	Wolfson Wing, KCL	SOCS[k]	0.2	25	^{13}C, ^{15}N	HCCCONH	Bruker 700 MHz
28/01/2013	Wolfson Wing, KCL	SOCS -EloBCM	0.2	25	^{15}N EloBC	^{15}N-HSQC for binding test	Bruker 700 MHz
28/02/2013	NMRC, Mill Hill	HIV2 SOCS	0.05	25	^{13}C, ^{15}N	HNCACB	Bruker 700 MHz

14/03/2013	Wolfson Wing, KCL	SOCS -EloBC[m]	0.19	25	^{15}N EloBC	^{15}N-HSQC for binding test	Bruker 700 MHz
		SOCS -EloBC	0.15	25	^{13}C, ^{15}N EloBC	^{13}C, ^{15}N-inter-NOESY	
11/04/2013	NMRC, Mill Hill	SOCS -EloBC[n]	0.1	25	^{13}C, ^{15}N EloBC	^{13}C-inter-NOESY	Varian 800 MHz

a. NMRC is the abbreviation of National Medical Research Centre. KCL is the abbreviation of Kings College London

b. 'SOCS -EloBC' stands for the complex of SOCS and EloBC. 'SOCS and EloBC' stands for two subunits separately according to experiments.

c. Two spectra, CC(CO)NH and H(CCO)NH.

d. NOESY was run at 4 ℃.

e. The sample is the same one run on 17/01/2011.

f. ^{13}CO-labeled proline and glutamine. Selectively-labeled EloBC in complex, 100% deuterated.

g. The same aliquot from sample 14/02/2012.

h. SOCS mutation III: P164S

i. The same aliquot from sample 11/05/2012

j. With two different mixing time: 8 ms and 17 ms.

k. The sample is the same one run on 20/12/2012. But the spectrum failed.

l. Three mutants: 101A, 102A and 103A.

m. Free EloB104 mutant was run first. Then SOCS -EloB104 complex.

n. 3/14 sample was changed to 100% D_2O buffer by using a sizing column and the experiment repeat using the same pulse-program on Varian.

Table 2.5 Summary of 3D spectra parameters (Continue)

Date	Spectrum	Protein Name	Pulse-program	Mixing time	Running time
27/11/2010	3D-NOESY	SOCS -EloBC	mh_wgnoesyhsqc_n15	90 ms	70 hrs
13/12/2010	3D-NOESY	SOCS -EloBC	/[a]	120 ms	67 hrs
17/01/2011	COSY	SOCS -EloBC	hcchcogp3d	60 ms	22 hrs
	TOCSY		hcchdigp3d	60 ms	22 hrs
	C -CCONH		hccconhgpwg3d3		66 hrs
	H-CCONH		hccconhgpwg3d2		40 hrs
20/01/2011	3D-NOESY	SOCS and EloBC	mh_wgnoesyhsqc_n15	90 ms	65 hrs
10/02/2011	^{15}N-TOCSY,	SOCS monomer	/	70 ms	44 hrs
	^{15}N-NOESY		/	90 ms	46 hrs
08/03/2011	HNHA	SOCS -EloBC	hnhaqg3d		40 hrs
31/03/2011	^{13}C-NOESY	SOCS -EloBC	mh_c13noesy_tf_4	90 ms	69 hrs
	HNCO		hncogp3d		15 hrs
22/05/2011	^{13}C-TOCSY	SOCS -EloBC	hcchdigp3d	?[b]	25 hrs
	^{13}C- NOESY		hesyhsqcetgp3d	120 ms	47 hrs

23/06/2011	^{15}N-NOESY	SOCS -EloBC	/	90 ms	65 hrs
21/07/2011	^{15}N-TOCSY	SOCS -EloBC	dipsihsqcf3gpso3d	60 ms	33 hrs
	^{15}N-NOESY		noesyhsqcf3gp193d	150 ms	58 hrs
19/01/2012	Trosy-^{13}C-HNCO	SOCS -EloBC	/		63 hrs
10/05/2012	^{15}N-TOCSY	MIII-EloBC	/	40 ms	34 hrs
	^{15}N-NOESY		/	120 ms	55 hrs
11/05/2012	^{15}N-TOCSY	SOCS MIII	dipsihsqcf3gpsi3d	?[b]	47 hrs
	HNHA		hnhagp3d		45 hrs
19/07/2012	HNCA	HIV2 SOCS	hncagpwg3d		63 hrs
30/08/2012	HNCA	HIV2 SOCS	/		60 hrs
30/10/2012	^{15}N-NOESY	HIV2 SOCS	mh_wgnoesyhsqc_n15	90 ms	46 hrs
08/11/2012	HNCOCA	HIV2 SOCS	hncacogpwg3d		91 hrs
20/12/2012	^{13}C-TOCSY	SOCS	/	8 ms	27 hrs
			/	17 ms	43 hrs
15/01/2013	HCCCONH	SOCS	hccconhgpwg3d3		20 hrs
25/01/2013	HCCCONH	SOCS	hccconhgp3d2		72 hrs
28/02/2013	HNCACB	HIV2 SOCS	mh_hncacbwg_2		92 hrs

14/03/2013	^{13}C-inter-NOESY	SOCS -EloBC	noesyhsqcgpwgx13d	100 ms	66 hrs
	^{15}N-inter-NOESY	SOCS -EloBC	Noesyhsqcf3gpwgx13d	100 ms	46 hrs
11/04/2013	^{13}C-inter-NOESY	SOCS -EloBC	/	125 ms	85 hrs

a. The pulse-programs from Varian machines are not listed here. All parameters were recorded in procpar file in each experiment.

b. The mixing time was not found in the log files (acqu or pulse-program files).

2.6.3 Paramagnetic relaxation enhancement studies

Gly-143, Asn-158 and Arg-167 of the SOCS-box peptide were mutated to Cys respectively in order to insert a Cys point-mutation for use in paramagnetic labelling. These three residues are selected for the reason that they are known to be outside two functional regions, the BC-box and the PPLPS motif, and have less biological functions. All mutants had been checked by ITC to verify that the mutation did not interfere with the binding affinity of SOCS-EloBC. The mutant protein was mixed with an approximately 5-fold excess of dithiothreitol (DTT) for two hours after elution from the Ni-NTA column. This was followed by separation of excess DTT by gel filtration chromatography. The SOCS-box monomer sample from the sizing exclusion column was collected and incubated with either the diamagnetic (1-acetyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulphonate or the paramagnetic (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-methyl) methanethiosulphonate (Toronto Research Chemicals) overnight at 4 ℃. The modified SOCS-box sample was dialyzed against NMR buffer and mixed with ^{15}N-labelled EloBC in NMR buffer at a 1.1:1 ratio. The mixed sample was then transferred to an NMR tube for NMR data acquisition. A ^{1}H-^{15}N HSQC spectrum was recorded for each 50 μM sample with three-hour acquisition. Intensity ratios were converted to distances according to an established model optimized from the Solomon–Bloembergen equation (Battiste and Wagner 2000).

2.7 Solution structure calculation and analysis

2.7.1 Fundamentals of structure calculation

As is detailed in the NMR theory part, each spin, i.e. ^{1}H, ^{15}N and ^{13}C in this work, has its particular physical character such as spin state, angular momentum, coupling frequency, resonance, etc, affected by the spin type and the surrounding micro-environment. These unique characteristics of a spin are displayed as chemical shifts and peak intensity in the spectra, and thus distances between ^{1}H spins are able to be acquired, from which the orders of a group of

residues in the solution can be calculated. As for the protein, by similar protocols, the orders of amino acid residues in a sequence are able to be determined.

Currently several tools are developed to determine NMR solution structures. The most common and typical way for a single macro-molecule to be solved is by using distance restraints from NOESY spectra, which has been widely used in the past decades of years (Banks, Hare *et al*. 1989, Nerdal, Hare *et al*. 1989). Briefly, under a NOESY pulse-program, there is a nuclear Overhauser effect (NOE) (Overhauser 1953) between two protons that are closer to each other in space within 5 Å. In a 3D spectrum, the resonance from one spin can be observed on the other spin and its strength is decreased according to the real distance between these two spins according to the principle that the strength is proportional to the inverse sixth power of the distance (Anderson and Freeman 1962). Each distance matches a pair of assigned protons between which a NOE transfer takes place through space rather than through bonds and is used as a distance restraint in structural calculations. Therefore, NMR solution structures can be determined from hundreds of these distance restraints.

Residual dipolar coupling (RDC) data is another common restraint in studies of biomolecular structures (Deloche and Samulski 1981, Tjandra, Garrett *et al*. 1997, Tjandra, Omichinski *et al*. 1997). This type of experiments is based on the idea that a molecule surrounded by liquid crystals such as lipid bicelles is aligned to an external magnetic field at a particular orientation, which provides a unique angle vector $\boldsymbol{\theta}$ of each bond such as N-H or C-H relative to the external magnetic field. This angle can be derived from the following formula:

$$D_{IS}(\theta) = \frac{\hbar \gamma_I \gamma_S}{4\pi^2 r_{IS}^3}[1 - 3\cos^2\theta] \tag{2.10}$$

where D_{IS} represents the residual coupling observed from spectra and r represents the distance between two spins. γ_I and γ_S are the gyromagnetic ratios of spin I and S respectively (Tolman, Flanagan *et al*. 1995). Differing from NOE distance restraints, RDCs provide global long-distance restraints considering that they provide relative orientations of some regions of a protein that are far apart in space. For example, in α-helix the angles of N-H bonds towards the external magnetic field are in theory the same whereas in an anti-parallel β-sheet angles on one sheet are opposite to those on the other sheet. One advantage of this technique is that it is less time-consuming compared to NOESY methods. The other advantage is that it does not require

the side-chain assignments, which prevents the errors from ambiguous NOESY data and poor-quality distance measurements. By this technique the relative orientations among long-distance separated residues can be acquired, thus determining the structures.

Site-specific labelling with paramagnetic tags is a newly-developed method providing greater distance restraints in recent years (Battiste and Wagner 2000). Briefly, the cysteine-connected chemical label, known as paramagnetic chemical, broadens peaks from residues closest to the label, which is called a paramagnetic relaxation enhancement effect. Its diamagnetic analogue, the control, does not generate such an effect. Similar to the NOE effect, the intensity is proportional to the inverse sixth power of the distance (Battiste and Wagner 2000). This phenomenon is observed between two spins up to 25 Å away. Another characteristic of this method is that only one cysteine is allowed in a protein sequence, which means that the observed effect happens exactly between an assigned residue and the labeled cysteine site. Due to these advantages, PRE is useful in studies of complexes (Korotkova, Yang *et al*. 2008). All these methods used to acquire long-distance restraints, including various pseudo-contact shift methods, have been well reviewed (Saio, Yokochi *et al*. 2010, Su and Otting 2010).

Besides the experimental-data-dependent structure calculations, structural modelling or prediction is becoming more advanced and reliable with the development of algorithms and computer clusters. One of the robust modelling software is the Chemical-Shift-ROSETTA (CS-Rosetta) (Shen, Lange *et al*. 2008, Shen, Vernon *et al*. 2009). This approach requires the input of the backbone assignments (NH, Hα, CA, CB, CO, N). In the first step, it performs data mining in the structural database where protein chemical shifts are deposited according to the assignment so as to acquire hundreds of fragments. These fragment candidates are used to assemble all-atom models by a Rosetta Monte Carlo protocol in the second step. The generated models are evaluated against the initial chemical shifts by SPARTA (Shifts Predicted from Analogy in Residue type and Torsion Angle) (Shen and Bax 2007). In the final step, all models are re-scored and aligned to the model with the lowest energy. If clustering around the lowest energy structure is converged, this run of calculations is regarded as ‘successful’. It is of note that distance restraints and RDC data are now combined with CS-Rosetta, which improves the accuracy of the modelling. However, the disadvantage of this method is that it requires large calculation capability, i.e. computer clusters and storage space. For a single processer, it takes

approximately 15 minutes to generate a structure and 50 days to finish a 5000-structure modelling.

2.7.2 Protein structure calculation

In this study, the SOCS-box peptide structure was generated by CS-Rosetta on the Biological Magnetic Resonance Bank (BMRB) CS-ROSETTA server (condor.bmrb.wisc.edu/bbee/rosetta/) and accepted according to the principles of CS-Rosetta (Shen, Vernon *et al*. 2009). The structure with the lowest energy was further refined according to the Rosetta refinement protocol (Ramelot, Raman *et al*. 2009). EloB and EloC structures were generated *de novo* as well on the server and accepted when the r.m.s.d. was less than 0.5 Å against the published crystal structure (PDB ID: 3DCG).

Vif SOCS-box was docked onto the EloBC dimer by a High Ambiguity Driven biomolecular DOCKing (HADDOCK) approach (Dominguez, Boelens *et al*. 2003, de Vries, van Dijk *et al*. 2007) on the WeNMR web server (de Vries, van Dijk *et al*. 2010). In the process of docking, 5 residues in SOCS-box (Gln-146, Ala-149, Leu-163, Pro-164, Ser-165), 4 residues in EloB (Asp-101, Val-102, Met-103, Lys-104) and 2 residues in EloC (Ala-82, Leu-86) were selected as active residues. The interfacial residues sitting between the SOCS-box proline-rich motif and the C-terminus of EloB were allowed to move fully at all stages. A file with distance restraints from PRE and the published crystal structure (PDB ID: 3DCG) are provided with coordinate files and the restraints were always enforced during the calculation. Two thousand initial structures of the complex were generated and the best two hundred structures were chosen for explicit solvent refinement. The cutoff of the cluster is 5 Å, 5 structures per cluster. Default parameters excluding the settings above were taken.

2.7.3 Structural refinement and analysis

The best structure of the complex from the top cluster generated by the HADDOCK modelling interface was used for the further refinement through the refinement interface on the

server. 20 structures with the lowest energy were chosen as the final structures. Pymol was used to analyze the structures and to generate structural figures (Schrodinger 2010). The quality of structure coordinates was determined by PROCHECK (Laskowski, Rullmannn *et al*. 1996) and iCing (Doreleijers, Sousa da Silva *et al*. 2012). Assignments *vs* coordinate validation was performed on BMRB server (http://deposit.bmrb.wisc.edu/).

2.8 Crystallography

2.8.1 Overview of X-ray crystallography

X-ray crystallography is the most widely used method to obtain structures of biological macromolecules. Recent developments enable X-ray crystallography to solve various sizes of small molecules as well as proteins ranging from less than 10 kDa to over hundreds of kDa. The first protein structure solved by X-ray crystallography was completed in 1958 (Kendrew, Bodo *et al*. 1958) and the biological molecular structure best known to the public is the DNA double-helix, published in 1953 (Watson and Crick 1953). Conclusively, compared to NMR spectrometry and cryo transmission electron microcopy, X-ray crystallography has several advantages:

1) Almost all sizes of proteins are able to be investigated by X-ray crystallography.
2) X-ray crystallography is the only technique that provides structural information at atomic resolution.
3) Crystal data obtained from a single solid state can be more accurate, whereas NMR usually offers an average solution structure from many conformational states.

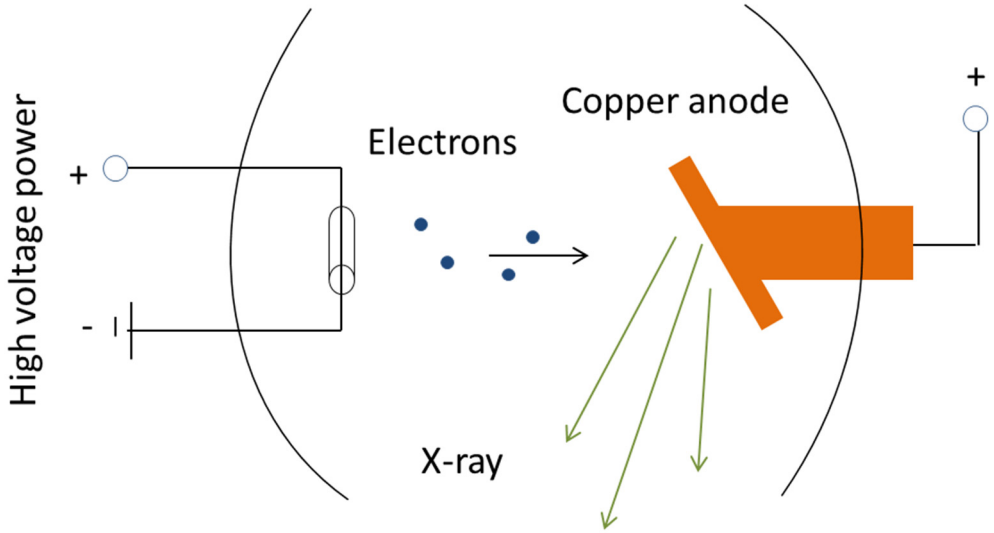


Figure 2.4: X-ray generation. Electron beams with high energy hit a copper anode, knocking out electrons from inner orbitals. Electrons in higher orbitals fall into the lower orbitals, releasing X-ray radiation.

X-ray radiation is a form of very short wavelength radiation (0.01-10 nm) generated from electron falling from a higher state to lower state. In order to produce X-rays, a high energy electron beam is accelerated using a high voltage to knock out electrons from inner orbitals of the atoms of the anode which is made of metals like zinc or copper. Electrons in higher orbitals, i.e. with higher energy, subsequently fall into the low-energy shells, releasing energy in the form of radiation. The radiation generated has a certain amount of X-ray quanta whose wavelength is dependent on the anode material (figure 2.4). Currently more X-ray sources from synchrotron radiation sources (SRS) are more used instead of in-house equipment because SRS is able to generate rays with less than 1 Å wavelength, which gives higher resolution data and improves the possibility of studying small crystals.

2.8.2 Fundamentals of crystallization

It is no doubt that in crystallography the first step, also one of the most critical and challenging steps, is to obtain a crystal of a high-quality in terms of its X-ray diffraction. The basic idea of growing crystals is to mix protein samples with a precipitant which is used for moving the protein sample to a supersaturated state in solution. In a certain kind of solution condition, there are three phases associated with increasing concentration of protein and precipitant respectively, which can be simply plotted on a phase diagram (figure 2.5). In the first phase, i.e. low amount of proteins and precipitants, proteins are undersaturated. After improving the concentration of proteins and precipitants, proteins go into the second phase called the metastable zone where crystals usually grow. In the third phase where there is a high concentration of proteins and precipitants, proteins aggregated and start precipitating. It is worth mentioning that in order to form a crystal, proteins generally enter phase III to nucleate before re-entering to phase II (figure 2.5). The nucleation state in phase III is therefore specifically termed the nucleation zone. The full path of crystal formation and growth is represented by C-shape arrow.

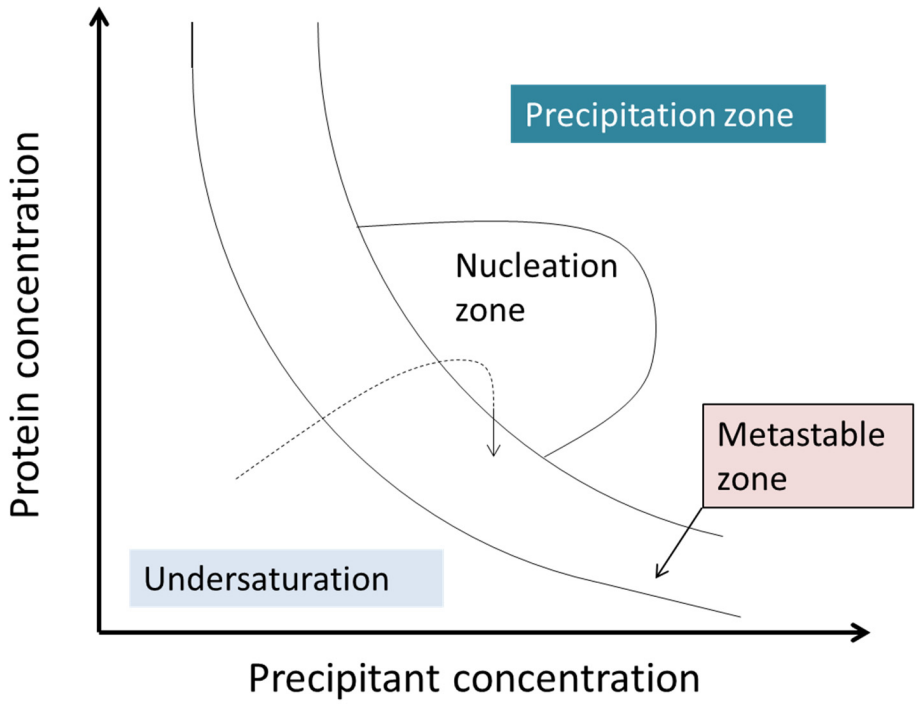


Figure 2.5: Protein phase diagram. The dotted line and arrow represent the pathway of forming crystals in solution for a molecule.

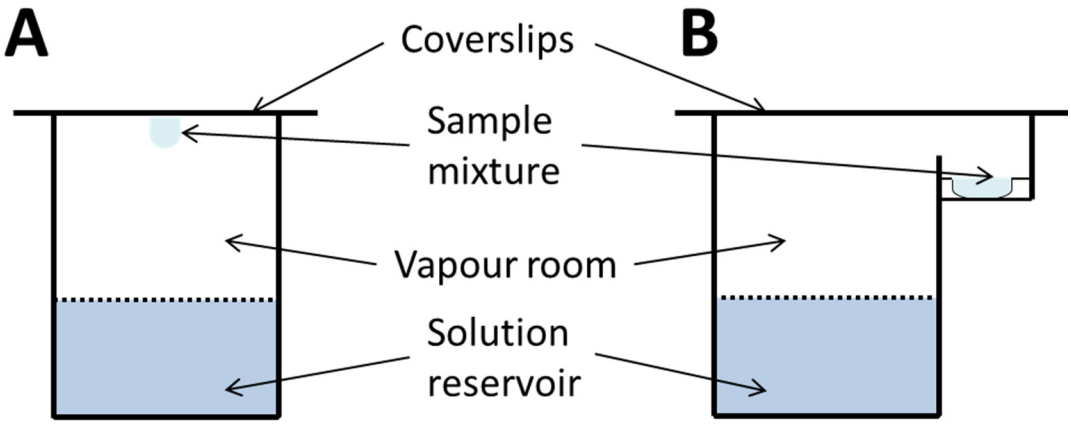


Figure 2.6: Vapour-diffusion equipment. Equilibration is established by vapouration in the room in each sealed reservoir. (A) Hanging drop method. (B) Sitting drop method.

To achieve crystallization, two methods are widely used based on vapour-diffusion principle. The first method is named 'hanging drop' (figure 2.6A), in which the mixture of protein and solution is added and hung on a siliconised coverslip. The other one is named 'sitting drop', in which the mixture is loaded into a well in a small sealed space (figure 2.6B). In both methods, the precipitating solution is equilibrated against a large amount of that in the reservoir through evaporation, during which the phase of the solution changes slowly.

A crystal can be described as a three-dimensional array where molecules are assembled in a very strict geometrical order. A crystal cell, or a unit cell, is viewed as the basic component of a crystal which has internal symmetry to other cells. The symmetry of a unit cell, or a space group, is revealed at a diffraction level as equal intensities of reflections in rotational symmetry and systematic absences in translational symmetry. There are six parameters to describe a crystal cell, **a**, **b** and **c** standing for three axes, and α, β and γ corresponding to angles between each axe (figure 2.7). Depending on the values of each parameter, a crystal lattice is classified as:

i. Cubic (**a** = **b** = **c**, α = β = γ = 90 °)
ii. Trigonal (**a** = **b** ≠ **c**, α = β = 90 °,γ = 120 °)
iii. Tetragonal (**a** = **b** ≠ **c**, α = β = γ = 90 °)
iv. Rhombohedral (**a** = **b** = **c**, α = β = γ ≠ 90 °)
v. Orthorhombic (**a** ≠ **b** ≠ **c**, α = β = γ = 90 °)
vi. Monoclinic (**a** ≠ **b** ≠ **c**, α = β = 90 °,γ ≠ 90 °)
vii. Triclinic (**a** ≠ **b** ≠ **c**, α ≠ β ≠ γ ≠ 90 °)

2.8.3 Fundamentals of X-ray diffraction and structure determination

Once a crystal is obtained and placed in the X-ray beam, it diffracts X-rays and thus generates diffraction according to the Bragg's law (figure 2.8):

$$2d \bullet \sin(\theta) = n\lambda \tag{2.11}$$

where θ is the angle of incidence and d is the distance between two planes of crystals.

Miller indices (h, k, l) are used to describe a crystal which is regarded as a series of

molecular planes. The three coordinates in Miller indices are defined as h = a/a′, k = b/b′, l = c/c′, where a′, b′, c′ are the coordinates at which the plane (h, k, l) intersects axes **a**, **b** and **c** respectively. By this, the reciprocal lattice vector for a certain plane is defined as:

$$G_{hkl} = ha^* + kb^* + lc^* \tag{2.12}$$

where a^*, b^* and c^* are the reciprocal axes unit vectors in the diffraction space from:

$$a^* = \frac{b \times c}{(abc)},\ b^* = \frac{c \times a}{(abc)},\ c^* = \frac{a \times b}{(abc)} \tag{2.13}$$

Therefore, when atom j within a unit cell is observed by X-ray beams and the phase defined as ***sr*** of an electron of the atom is an integer, the phase differences between atoms are 2nπ and the scattering vector s of atom j is defined as:

$$r_j \bullet s = asx_j + bsy_j + csz_j = 2\pi(hx_j + ky_j + lz_j) \tag{2.14}$$

where r_j is the coordinate of individual electrons, s is the vector of scattering wave and x_j, y_j, z_j are the fractional coordinates of atom j. If the electron density is continuous throughout a unit cell, a factor for an atom j called the 'crystallographic structure factor' can be driven from integral equation:

$$F_{hkl} = \int_V \rho(r) \exp(2\pi i r \square s) dv \tag{2.15}$$

in which $\rho(r)$ is the electron density in the space at position r. Subsequently, the electron density can be calculated from the structure factor by an inverse Fourier transformation:

$$\rho(r) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i r \square s) \tag{2.16}$$

where V is the volume of a unit cell. By applying the Friedel's law that intensity of a reflection is equal to that of its centro symmetric related reflection, i.e. $I(h, k, l) = I(-h, -k, -l)$ mathematically and replacing $r*s$ with the equation (14), equation (16) can be substituted to:

$$\rho(r) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| \cos(2\pi(hx_j + ky_j + lz_j) - \phi_{hkl}) \tag{2.17}$$

In the equation (17) the new parameter Φ_{hkl} is the phase for the structure factor $|F|_{(hkl)}$ amplitude and cannot be acquired directly from experiments, which is known as the phase problem in crystallography. After the phase is determined, equation (17) is finally applied to calculate an electron density map, where the electron density indicates positions of atoms and the height of

density is proportional to the number of electrons in an atom.

In order to address the phase problem and thus to solve the structure, there are several methods available. Experimentally, if a crystal is co-crystallized with some heavy atoms that occupy specific positions within the unit cell without disturbing original structures, the structure factor differences between data sets collected from native or derivative samples can be used to establish the positions of the heavy atoms by using known heavy atom positions (Green, Ingram *et al*. 1954). Similarly, the single anomalous diffraction method is another alternative to localize some specific atoms (Rice, Earnest *et al*. 2000). This idea is based on the fact that at certain wavelengths some atoms diffract X-rays anomalously. Positions of these atoms are acquired from the collected data sets are thus used to restore the phases.

Besides the experimental methods, Molecular Replacement method (MR) is a widely-used way to address the phase problem using only the originally collected data sets. Briefly, previously-solved structures of homologous macromolecules are used as starting models. In the first step, the starting model is used to calculate structure factors and phases during each replacement, which are then compared with experimental structure factors. In the second step, structure factors, including observed and calculated factors, and the model phases are applied in mapping the electron density that, in turn, is used to correct the generated model. This process of correcting models is called refinement which is repeated until calculated structure factors have a satisfactory agreement with experimental structural factors. Detailed information on MR has been reviewed by Delarue (Delarue 2008).

The resolution of a model is an important criterion that decides what kind of information can be obtained from crystals and diffraction. Usually a larger and more ordered crystal is able to provide higher resolution, and as a result the composition of a unit cell can be observed more clearly. Basically, around 6 Å resolution only suggests the arrangement of α-helices. Positions of side-chains are resolved at 3 Å resolutions whereas the position of each atom can only be determined at 2 Å resolution or higher.

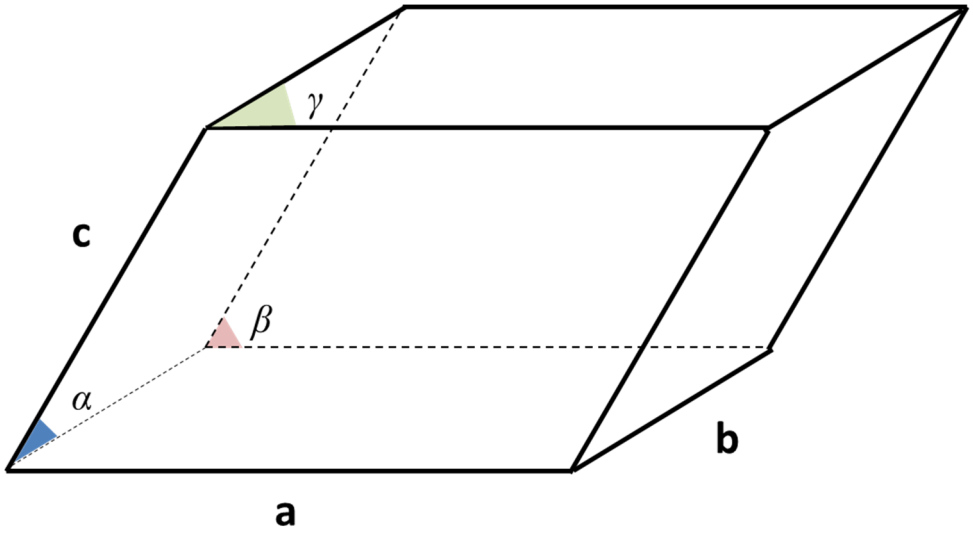


Figure 2.7: Six parameters of a crystal cell.

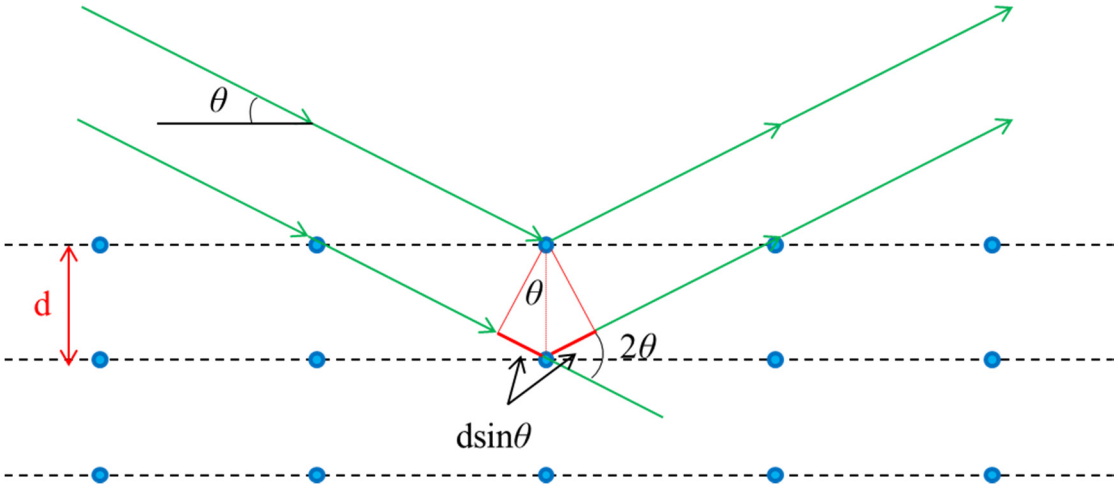


Figure 2.8: The diagram of Bragg's law. The incident angle is θ and the distance between two crystal planes is d. The path difference between two light paths is highlighted in red, which is 2dsinθ.

2.8.4 Vif/EloBC/CBFβ complex crystallization

The sample of Vif/EloBC/CBFβ complex was concentrated to 4 mg/ml. As for high-throughput condition screening (HTS) in Hauptman Woodward Medical Research Institute (HWI) (Buffalo, New York, U.S.A.), 700 µl sample was prepared each time and shipped in a thermos covered by ice bags to the lab, which usually took four days to arrive. Various types of crystallization kits were tested in 96-well sitting-drop MRC plates (Molecular Dimensions) set by Mosquito® in-house. A 0.6 µl protein sample was mixed with 0.6 µl and 0.3 µl crystallization solution respectively. As for hanging-drop scanning, 500 µl solution was added in the reservoir of a 24-well Linbro-style vapour diffusion plate (Molecular Dimensions). Protein and solution was mixed on a siliconized cover slide (Jena Bioscience) which was sealed onto the reservoir. All plates set up in-house were placed in a cold room (4 ℃). Crystals were checked for diffraction by using a PX-scanner manufactured by Agilent.

2.9 Biological assays

2.9.1 Immunoblot analysis

To test the Vif expression and purification from *E. coli*, 2 µl of 10-fold diluted samples in each purification step was added onto a 15% acrylamide gel. For the western blotting from cellular assays, 15 µl of each sample was loaded onto 15% acrylamide gel. Gel was run at 120 volts for 2 hours. Afterwards proteins were transferred onto a nitrocellulose membrane at 16 V overnight in transfer buffer (20% methanol, 0.2 M glycine, 0.1% SDS, 27.6 mM Tris). The membrane was blocked in blocking buffer (2.7 mM KCl, 137 mM NaCl, 2 mM KH_2PO_4, 10 mM Na_2HPO_4, pH 7.4, 0.5% Tween-20, 2% milk) for one hour, and then incubated with appropriate primary antibodies diluted to the acquired concentration in PBST buffer (2.7 mM KCl, 137 mM NaCl, 2 mM KH_2PO_4, 10 mM Na_2HPO_4, pH 7.4, 0.5% Tween 20) for one hour or overnight. After washing with PBST buffer three times, three minutes each, membranes were incubated with appropriate HRP-conjugated secondary antibodies for one hour. After three

washes of 5 min each in PBST buffer, proteins were detected by Thermo Scientific® Pierce* ECL Plus western blotting substrate and visualized using the Odyssey® imaging system (LI-COR).

2.9.2 Cell culture and transfection

293T cells derived from human embryonic kidney fibroblasts were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 1% penicillin and 10% fetal bovine at 37 ℃ in 5% CO_2. Cells are dealt with TripLE™ and split every two days.

To transfect DNA into 293T cells, the appropriate amount of DNA was mixed with 4 ug of polyethylenimine (PEI, working stock concentration 1 mg/ml) per ug of DNA diluted in 500 µl Opti-MEM, and mixed by vortexing. After 20 minutes, the mix was added gently drop-by-drop to cells seeded at approximately $1x10^5$ in 6-well plates 24 hours before. Transfected cells were incubated at 37 ℃ in 5% CO_2 for 20 hours, after which the medium was changed. Transfected cells were usually harvested 48 h post transfection. All chemicals above were provided by GIBCO.

2.9.3 Infectivity assay

To produce infectious HIV-1 virions, 0.4 µg p8.91 (HIV-1 Gag-Pol), 0.4 µg pCSGW, 0.6 µg pMDG (G protein of vesicular stomatitis virus), 0.5 µg pA3G-hemagglutinin epitope (HA) or pCMV-HA (as a negative control for APOBEC3G) and 0.12 µg pVif from NL4.3 or mutants were co-transfected into 293T cells. The total amount of DNA was kept constant by adding plasmid pcDNA3.1, an empty parental vector. The virus-containing media was harvested 24 hours later, and filtered using a 0.45 µm filter (Milipore). Virus-producing cells were resuspended in 200 µl RIPA buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 2% NP40, 0.5% sodium deoxychloride) and then sonicated for 10 seconds. 10 µl of the sonicated cell lysis was taken and mixed with gel loading buffer for western blot analysis.

Viral infectivity was measured by challenging 293T cells seeded at around $1x10^5$ per well the day before infection in 24-well plates with 100 µl of a three-fold dilution series of the viral

stock. 10^5 cells per well were plated in 24-well plates 24 hours before infection. After 72-hour infection at 37 ℃ the media was removed and the cells were trypsinised and resuspended in PBST buffer with 2% formic acid. Green fluorescent protein $(GFP)^+$ expression cells were quantified by flow cytometry using a Calibur FACS machine (B&D).

2.9.4 Co-immunoprecipitation assay

293T cells were co-transfected with 2 μg pCul5-HA or pCMV-HA as a negative control, and 2 μg pVif or mutants. 2 μg pcDNA3.1 was used to replace pVif in Vif- groups. Transfected cells were incubated as described above, after which the medium was changed. Another 24 hours later cells were resuspended in 250 μl Co-IP buffer (150 mM NaCl, 50 mM Tris pH 7.0, 0.5% Triton X-100, Roche protease inhibitor tablet) and debris was removed by centrifugation at 6,000g for 1 min. 25 μl of the supernatant was mixed with 25 μl gel loading buffer and boiled for 5 min for western blot. The remaining supernatant was incubated with 30 μl buffer-balanced protein G-conjugated agarose beads (Invitrogen) for 3 hours at 4 ℃. The beads were collected by centrifugation at 10,000 g for 30 sec, and the supernatant was removed. After four washes with 1 ml co-immunoprecipitation (Co-IP) buffer for 5 min at 4 ℃, samples were separated from beads by elution with 30 μl Co-IP buffer containing HA peptides. Fractions were mixed with 30 μl loading buffer and boiled for 5 min. Both cell lysates and elutions were analysed by western blot.

CHAPTER 3: NMR STUDY OF HIV-1 VIF SOCS-BOX

As described in Chapter 1, the structure of Vif is still undetermined due to the problems of expressing soluble folded Vif. Our previous work (Bergeron, Huthoff *et al*. 2010) succeeded in acquiring a soluble Vif fragment including the critical BC-box and the proline-rich motif, which allows us to determine the function of Vif SOCS-box from a structural biological standpoint. NMR structural calculations were subsequently applied to reveal the structure of the Vif C-terminus 139-173.

3.1 NMR spectra of Vif SOCS-box peptide

3.1.1 HSQC spectra of the SOCS-box

The expressed protein is composed of two portions, a solubility-enhancement-tag (SET-tag) and the target Vif SOCS-box domain fused with a hexa-histidine tag at the C-terminus (figure 3.1A, residues 63-95 are corresponding to residues 143-175 in the Vif sequence). The ^{1}H-^{15}N HSQC recorded on free SOCS-box peptide reveals that the SET-tag is folded properly in the solution with a very good peak distribution whereas SOCS-box is unstructured (figure 3.1B). During the binding to EloBC, the peaks from residues 143-153 that are known to lie on the binding surface start broadening in the HSQC spectrum due to the decreased exchange rate (figure 3.1B, red spectrum). It implies that a conformational change of this domain undergoes. This perturbation provides important information in structural determination and whereas difficulties in assigning the SOCS-box peptide in bound state.

3.1.2 Towards the re-appearance of missing peaks in the bound state

In order to observe and assign more peaks for the bound SOCS-box peptide, ^{15}N-labeled SOCS-box in the SOCS-EloBC complex was prepared and tested over a temperature range

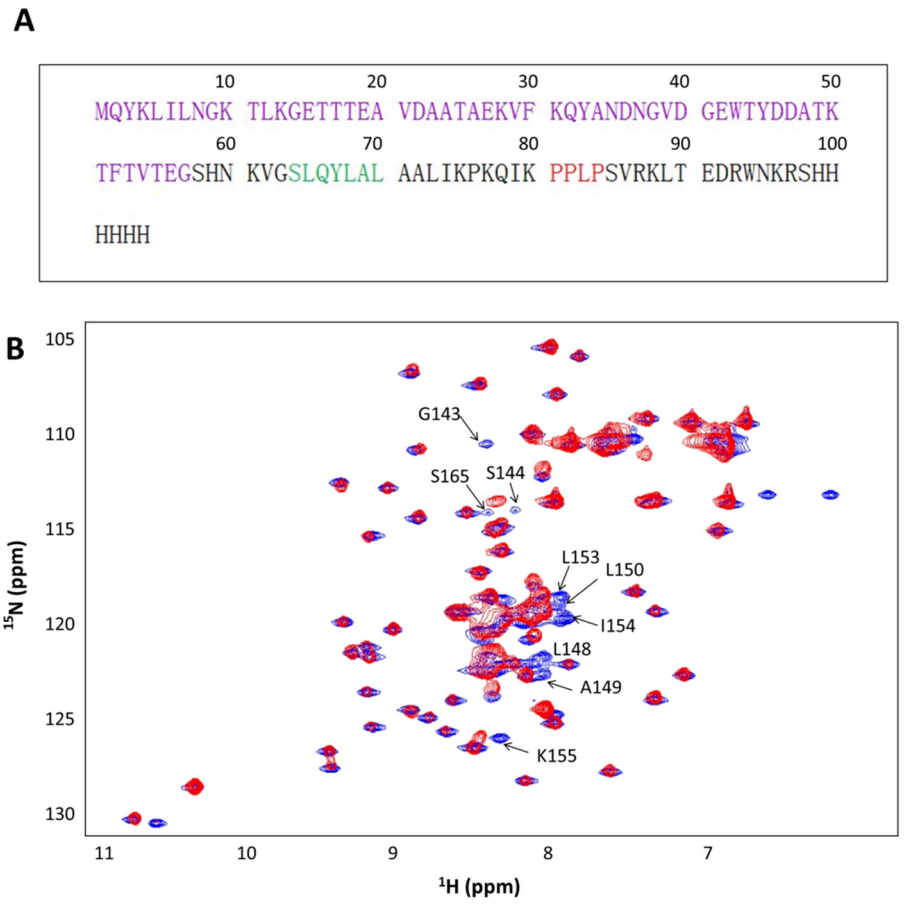
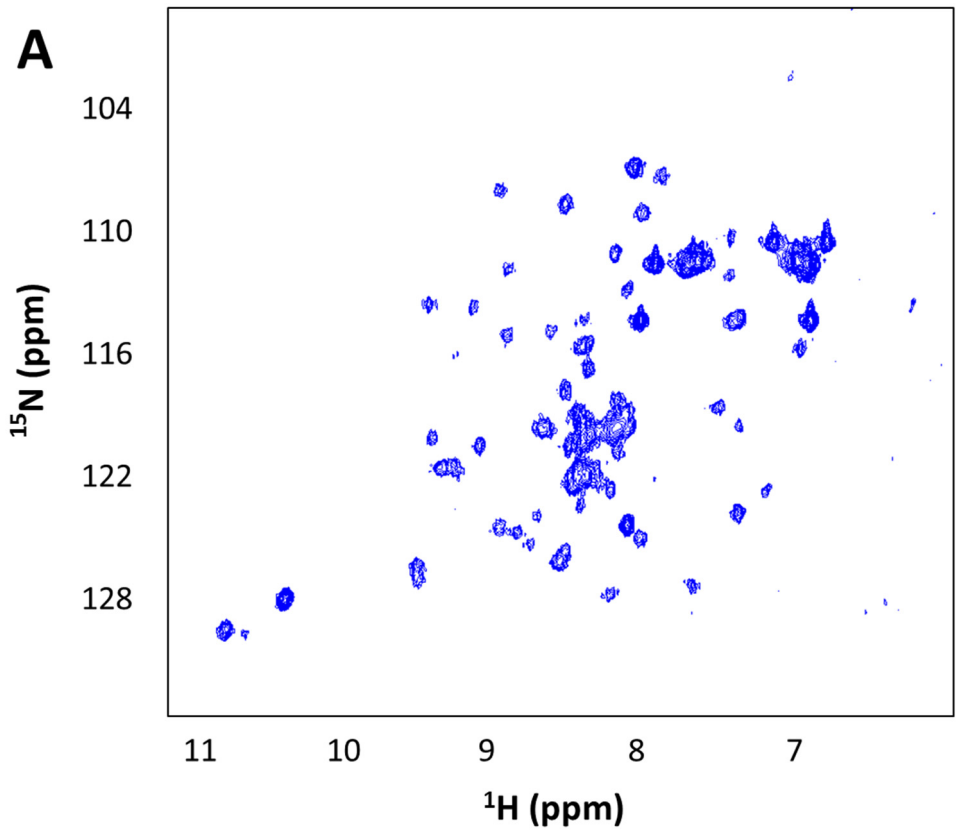
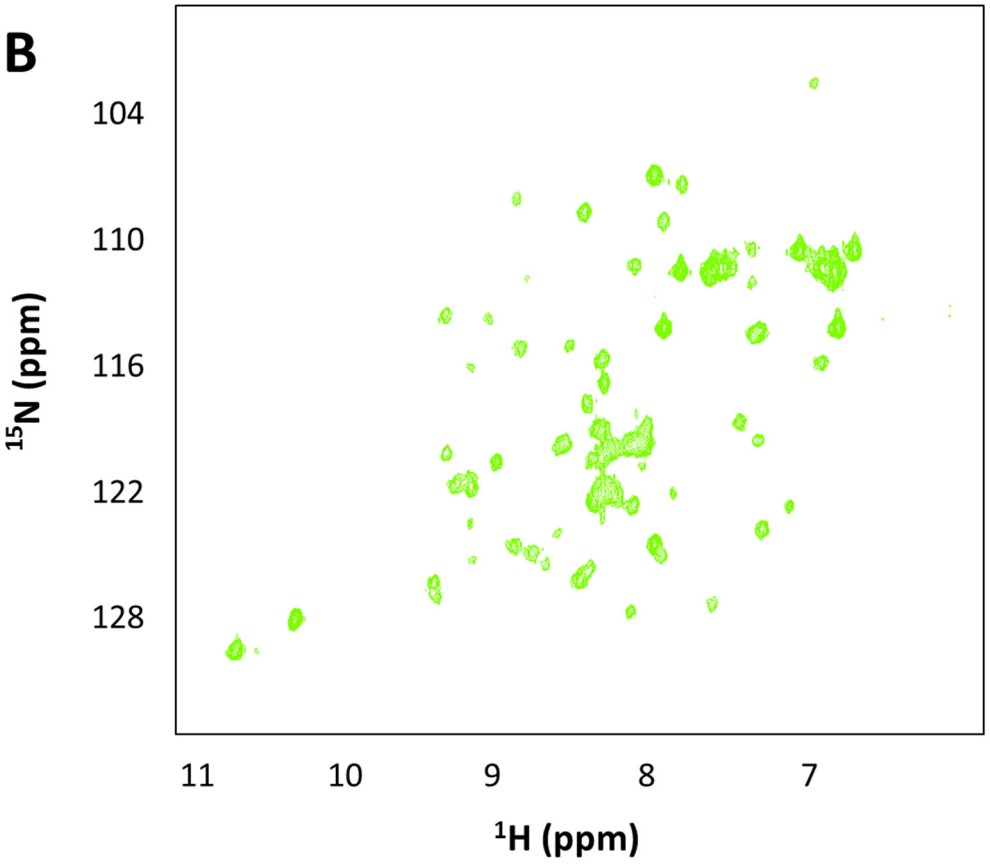


Figure 3.1: (A) The full sequence of the SOCS-box peptide fused with a SET-tag sequence coloured in violet. The BC-box in Vif is presented in green and the proline-rich motif is highlighted in red. Residues 63-95 in the expression peptide are corresponding to residues 143-175 in the full-length Vif sequence (B) Two ^{1}H-^{15}N HSQC spectra of the free SOCS-box (blue) and the bound SOCS-box peptide are overlaid. Missing peaks corresponding to SOCS active residues are indicated. These peaks were missing during the binding in NMR perturbation experiments. Reproduced from the reference: The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex (Bergeron, Huthoff *et al*. 2010).


A
104
110
116
122
128
^{15}N (ppm)
11
10
9
8
7
^{1}H (ppm)


B
104
110
116
122
128
^{15}N (ppm)
11
10
9
8
7
^{1}H (ppm)

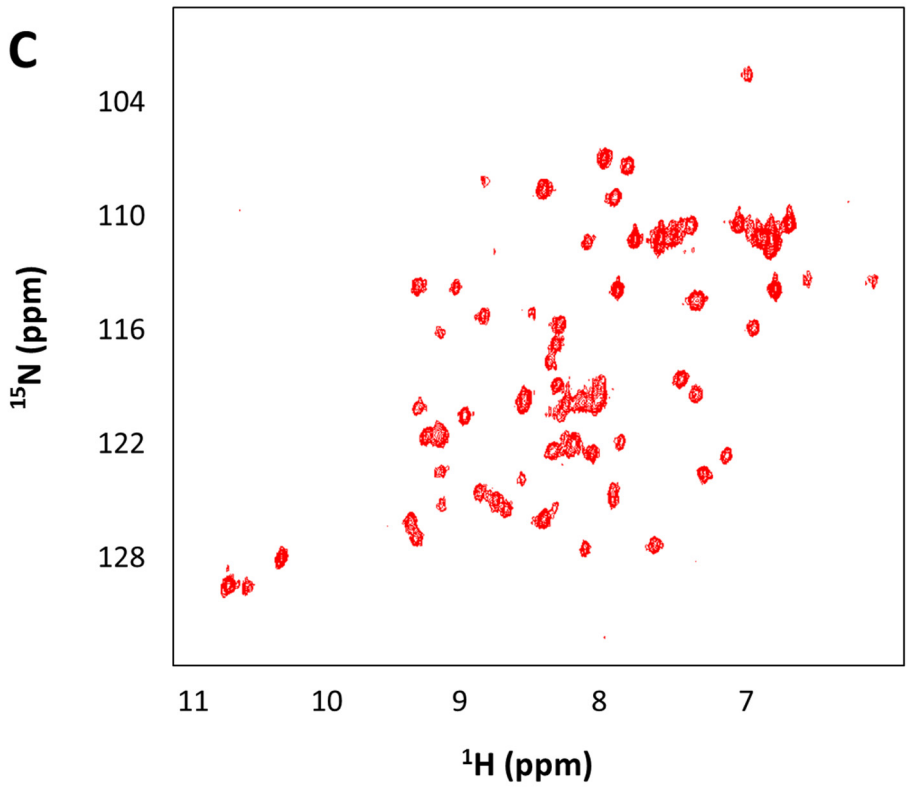


Figure 3.2: ^{1}H-^{15}N HSQC spectra of bound SOCS-box peptide was run at different temperatures: 25 ℃ (blue), 30 ℃ (green) and 37 ℃ (red) respectively.

from 25 ℃ to 37 ℃ (figure 3.2). More peaks in the central region of the spectrum are broadened with increasing of temperature. At 37 ℃ only a few peaks from the SOCS-box can be observed. Providing that Vif naturally functions at body temperature, and that EloBC spectra were recorded at 30 ℃, lower temperatures were not further tried. A small scale pH study (from pH 6.5 to pH 8.0) was performed, which did not induce any reappearance of peaks (data not shown).

3.2 SOCS-box assignment

3.2.1 Side-chain assignments

In normal assignment strategies, side-chains are able to be assigned after backbone assignments (CO, Cα, Cβ, NH and N) are confirmed. 3D spectra such as HCCH-COSY and HCCH-TOCSYs with a range of mixing time are usually applied to fully assign the side-chains, followed by ^{15}N-TOCSY-HSQC and HCC(CO)NH for validation. In this study we first ran a ^{15}N-TOCSY-HSQC, HCCH-TOCSY with 8 ms mixing time and HCCH-TOCSY with 17 ms mixing time. Usually ^{15}N-TOCSY-HSQC does not give enough side-chain information due to the large-distance connection between detected scalar coupled spins (figure 3.3A). A HCCH-TOCSY with a shorter mixing time usually provides unambiguous and secure information on neighbouring protons (figure 3.3B), and a longer mixing time offers more information on distinct protons for long side-chain groups (figure 3.3C). For each spectrum peaks were picked and assigned manually, using the software CcpNmr suite.

In the next step we applied HCC(CO)NH pulse program not only for the validation but also for the proline assignment in our case. Due to the cyclized side-chain proline cannot be observed in ^{15}N spectra. Therefore, before assigning the peaks from prolines in ^{13}C spectra, one starts with a limited number of proline assignments. Through a HCC(CO)NH experiment, some proton chemical shifts of a proline (i) were observed and determined from the next residue (i+1). By this procedure, most of the residues were assigned regardless of their presence in a proline-rich motif, improving the percentage of assigned backbone to 83.7% and that of assigned proton to 78%. Over 90% of resides were either fully or partially assigned. All the assignments

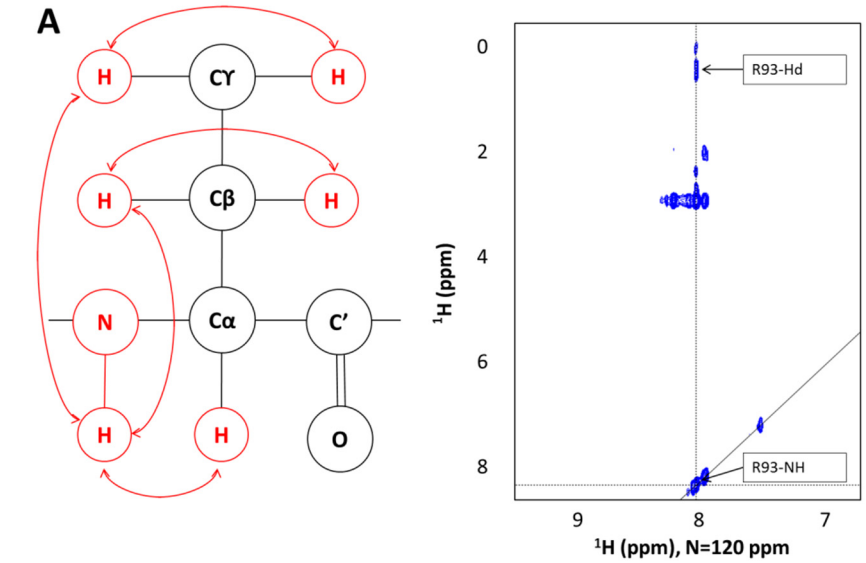
excluding the assignments on SET-tag are listed in Appendix III

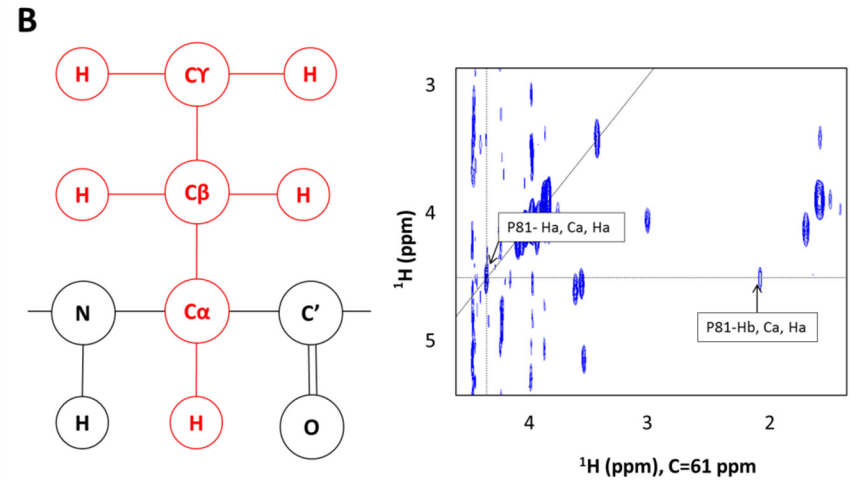
3.2.2 Nuclear Overhauser effect spectrometry assignment

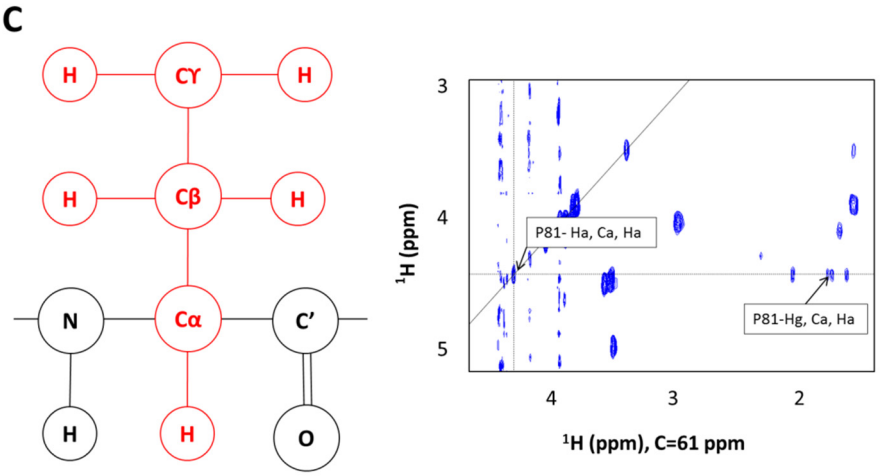
After side-chains had been assigned, the next step was to run 3D NOESY spectra (figure 3.3 D, E) to obtain NOEs and hence distance restraints for structural calculation. Peaks were picked and assigned manually using CcpNmr suite. However, instead of having peaks from the SOCS-box, most of the cross peaks were from the well-structured SET-tag, and some changes in the intensities of resonances were from two spins in the tag and the SOCS-box respectively. This is because of the fact that in a structured region a spin, i.e. a proton, has a large population of neighbouring spins through space and changes in the intensities of two resonances take place more possibly, thus more distance restraints can be obtained. On the contrary, in a flexible region most NOEs are from an observed residue (i) and its neighbour residues (i±1), which means that the number of NOEs is usually limited.

3.2.3 An approach to assigning the bound SOCS-box

Considering that SOCS-box experiences a structural change during the binding to EloBC, we chose to use backbone assignments from bound SOCS-box for CS-Rosetta structural calculation. ^{15}N-HSQC, ^{13}C-HSQC, ^{15}N-TOCSY-HSQC and HCCH-TOCSY spectra were run and processed. In the first step all peaks from SOCS-box in these spectra were picked. In the next step our assignments from free SOCS to the bound SOCS-box peptide were taken, which was able to provide assignments on HN, N, Cα, Cβ and Hα for CS-Rosetta calculation. In the final step ^{15}N-editted NOESY and ^{13}C-editted NOESY spectra were collected in order to measure more NOEs for distance restraints. It should be noted that as the structure of Vif BC-box in bound state had been solved by X-ray crystallography (Stanley, Ehrlich *et al*. 2008), it facilitated NOE assignment and improved the accuracy in spite of missing peaks from BC-helix domain. But due to the reasons elaborated above, only 78 unambiguous intra-molecular NOEs were acquired in our case.


A
H
Cγ
H
H
Cβ
H
N
Cα
C′
H
H
O
R93-Hd
R93-NH
0
2
4
6
8
9
8
7
¹H (ppm)
¹H (ppm), N=120 ppm


B
H
Cγ
H
H
Cβ
H
N
Cα
C′
H
H
O
P81- Ha, Ca, Ha
P81-Hb, Ca, Ha
3
4
5
4
3
2
¹H (ppm)
¹H (ppm), C=61 ppm


C
H
Cγ
H
H
Cβ
H
N
Cα
C′
H
H
O
P81- Ha, Ca, Ha
P81-Hg, Ca, Ha
3
4
5
4
3
2
¹H (ppm)
¹H (ppm), C=61 ppm

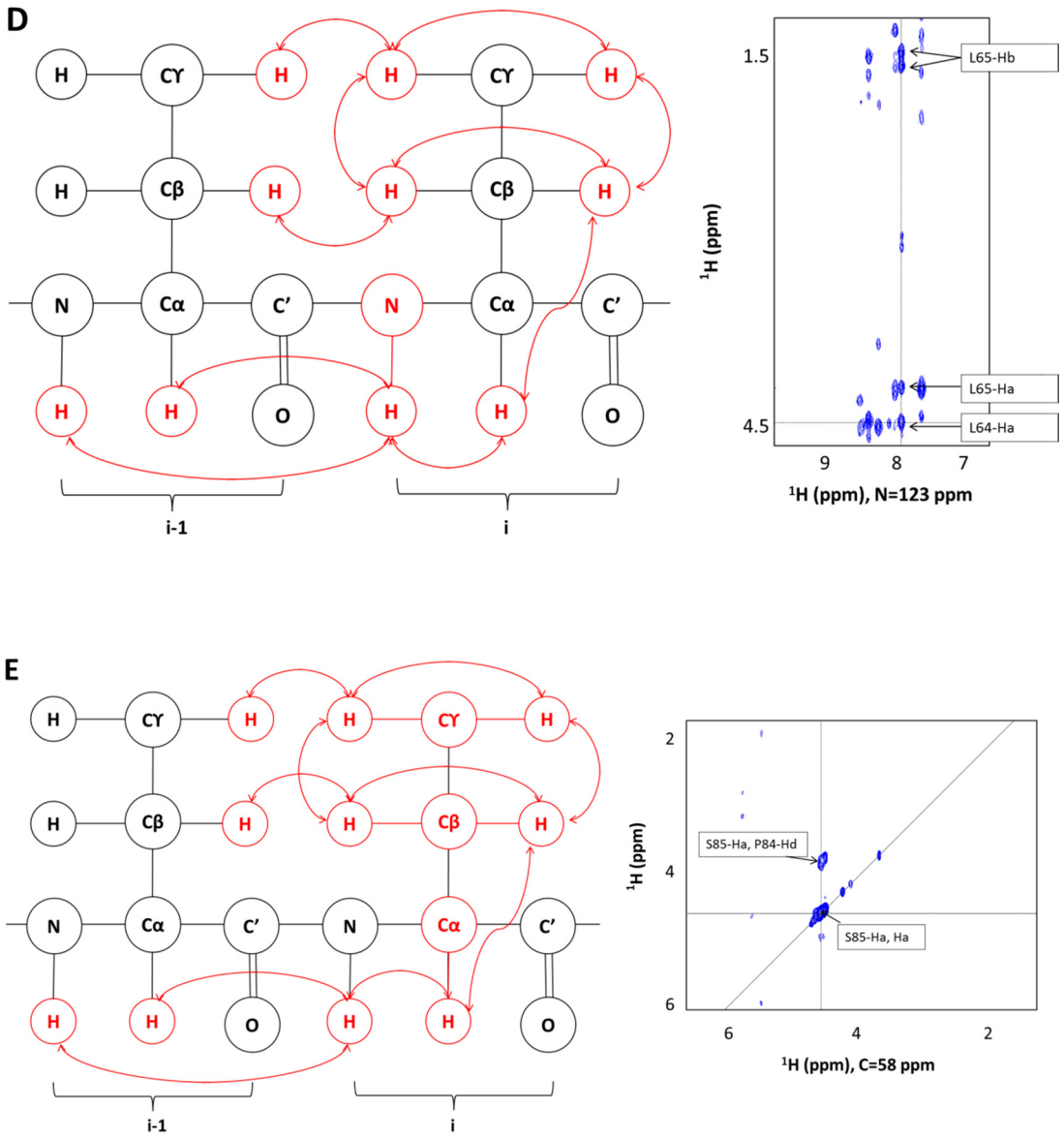


Figure 3.3: Explanation of various 3D spectra of the SOCS-box. For all spectra, the spin system and magnetization transfer coloured in red is presented on the left. Red arrows stands for 'through space'. A strip from the relative spectrum is selected and presented on the right. The assigned peaks are indicated. (A) ^{15}N-TOCSY-HSQC. (B) HCCH-TOCSY with 8 ms mixing time. (C) HCCH-TOCSY with 17 ms mixing time, which gives more information on sidechains. (D) ^{15}N-NOESY-HSQC with 120 ms mixing time. (E) ^{13}C-NOESY-HSQC. NOESY spectra provide information from other nearby residues.

3.3 Bioinformatic analysis

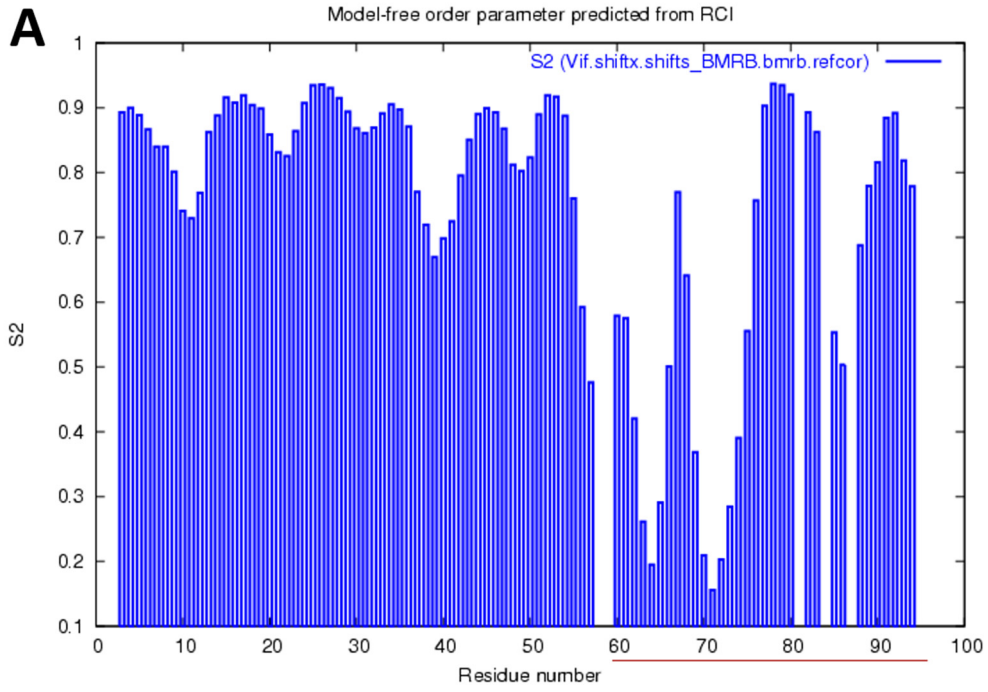
3.3.1 Flexibility analysis

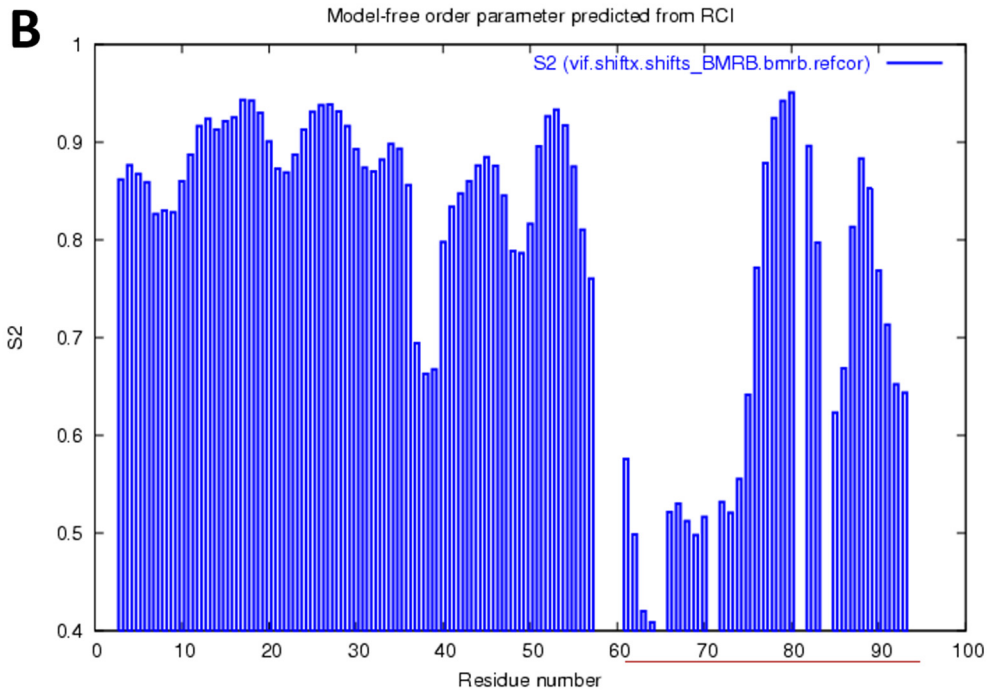
Although using flexibility predication software cannot provide much accurate structural information compared to other bioinformatic analysis, it gives a direct and general view on the behaviour of a protein in solution. In this study chemical shifts were required to identify flexible loops according to the protocol of CS-Rosetta (Shen, Vernon *et al*. 2009).

Our RCI analysis obtained for the SET-tag domain agrees with the published structure of the tag (Gronenborn, Filpula *et al*. 1991), showing that SET-tag (residue No. 1-58 in this construct) is completely structured in (figure 3.4A,B). The low RCI value of SOCS-box α-helix is caused by missing chemical shifts, which means the result for this domain may not be correct. Interestingly, RCI predicts that sequence 76-83 which includes residues 81PPL83 and sequence 87-91 is less disordered both in bound and unbound state (S2 > 0.7) as Dr Julien Burgeron previously did (Dr Julien Burgeron's thesis, 2010). This result, combined with a previous mutational study (Simon, Sheehy *et al*. 1999), may imply an important loop formed by 84PS85 in Vif function.

3.3.2 Torsion angle prediction

Being a part of the preparation of input data for CS-Rosetta, the role of TALOS+ is to identify the disordered tails and loops, which is similar to what RCI does. Following the result from RCI, TALOS+ also nicely predicts the angles of SET-tag (figure 3.4C). As for the SOCS-box, owing to the lack of sufficient chemical shifts TALOS+ fails to perform a prediction on the α-helical domain, the PPLPS motif and the carboxyl trail. The rest of the SOCS region is predicted well. The predicted angles by TALOS+ may imply a β-strand-like structure. At this point, according to the results from RCI and TALOS+, the C-terminus of SOCS-box (residue 174-177) was truncated during the structure generation.


A
Model-free order parameter predicted from RCI
S2 (Vif.shiftx.shifts_BMRB.bmrb.refcor)
S2
Residue number
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
70
80
90
100


B
Model-free order parameter predicted from RCI
S2 (vif.shiftx.shifts_BMRB.bmrb.refcor)
S2
Residue number
1
0.9
0.8
0.7
0.6
0.5
0.4
0
10
20
30
40
50
60
70
80
90
100

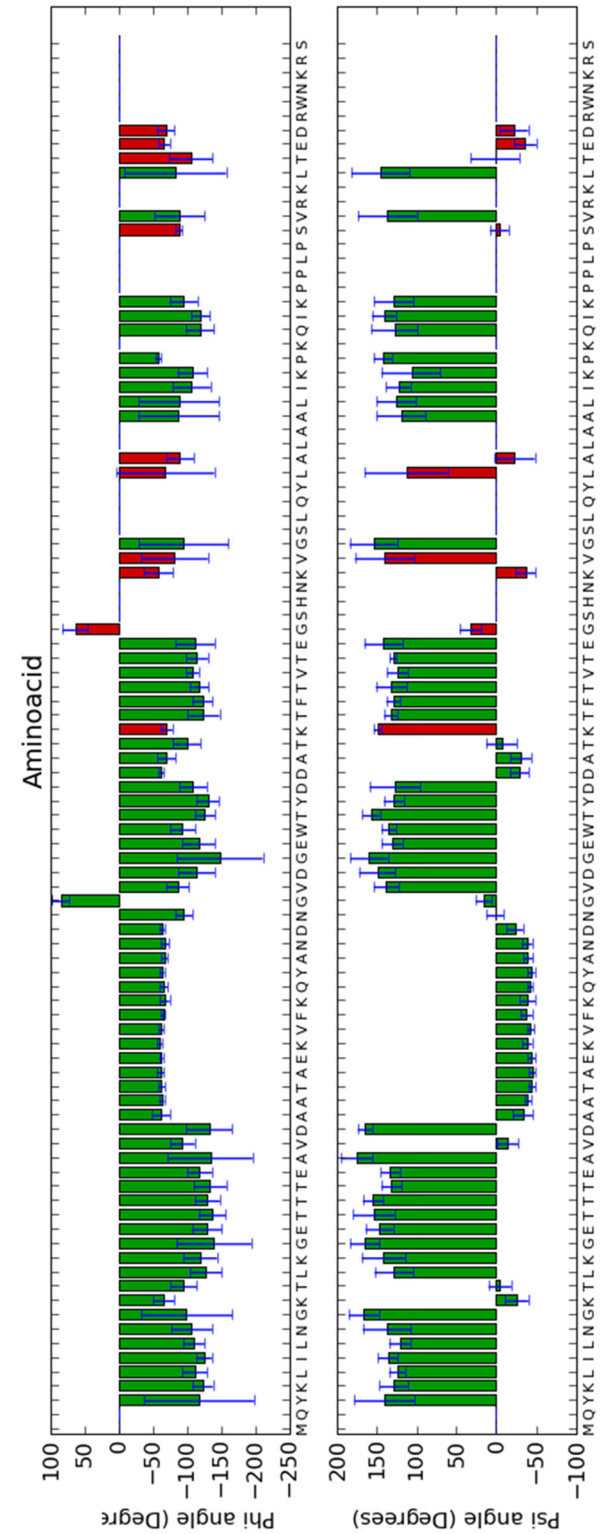


Figure 3.4: Analysis of SOCS-box amino acid sequence. (A) The flexibility prediction on the free peptide by RCI. The higher the value is, the less flexible the sequence is. (B) The flexibility prediction on bound peptide by RCI. The SOCS-box sequence is underscored. (C) The torsion angle prediction on bound peptide by TALOS. Green colour represents a good prediction and red colour represents a bad prediction.

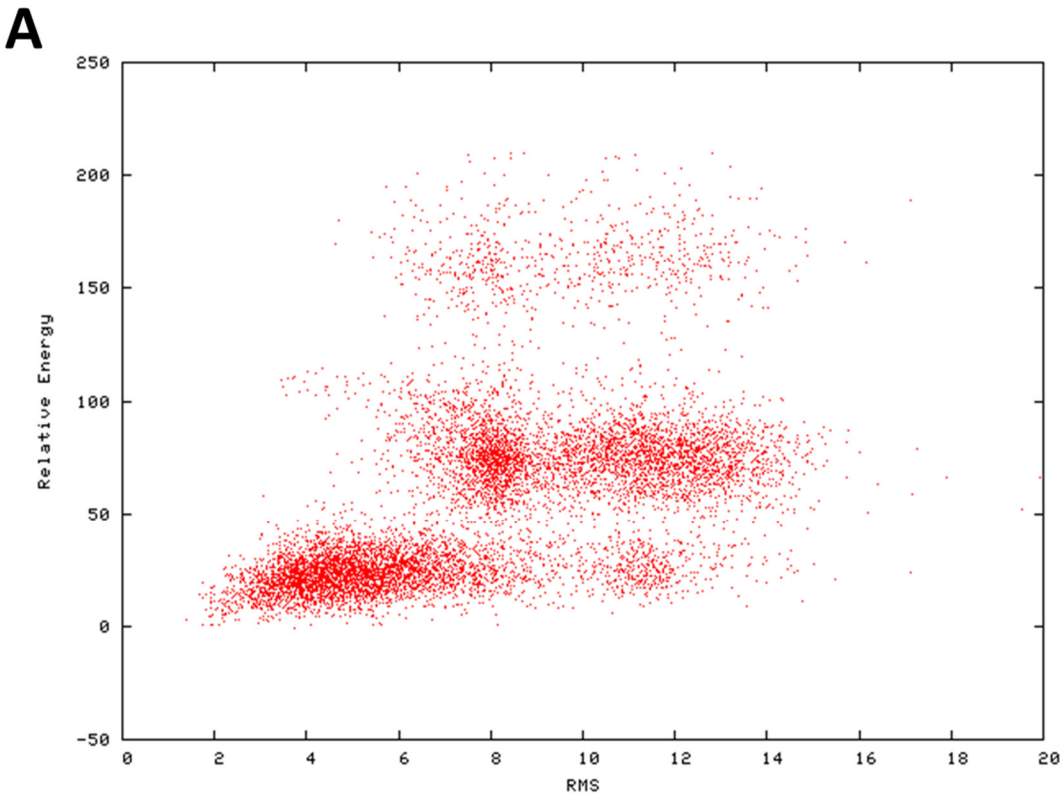
3.4 Solution structure calculation of Vif SOCS-box

Although most of the assignments from the α-helical motif of SOCS-box are unavailable due to peak broadening caused by the exchange rate, it is still reasonable to generate solution structures by CS-Rosetta (figure 3.5A). This is because the structure of sequence 163-173 has been determined by X-ray crystallography (PDB ID: 3DCG) and this published structure could be referred to in data mining stage of CS-Rosetta (Stanley, Ehrlich *et al*. 2008).

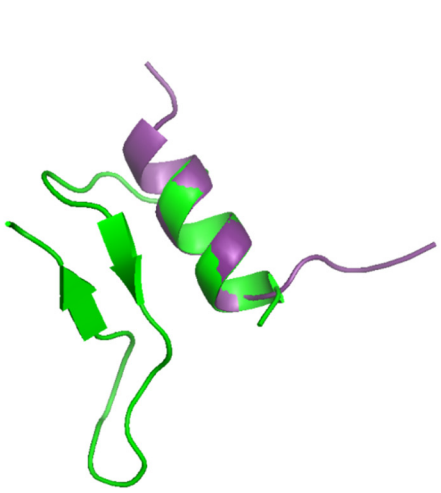
Superimposition of the refined ten lowest-energy models shows that a loop between two β-sheet is disordered and flexible (figure 3.5B) as it is predicted by RCI, and that the rest of sequences are structured and stable in the complex. The coordinates with the lowest-energy are aligned to the crystal structure, giving an r.m.s.d. at 0.411 Å (figure 3.5C), which means the structures slightly differ between the solid and solution states. The structural statistics for SOCS-box is given in table 3.1. It is worth mentioning that the chemical shifts from the β-sheet are not disturbed by the exchange rate, which may indicate the β-sheet does not exist on the interface of the interaction.

Table 3.1. Structural statistics for the SOCS-box in bound state

NOE restraints of SOCS-domain peptide	
Short-range	59
Medium-range	11
Long-range	8
Average RMSDs from the mean structure	
Backbone average (Å)	0.54 ± 0.19
Heavy atom average (Å)	0.65 ± 0.22
Ramachandran plot	
Most favoured region (%)	96.7
Allowed region (%)	3.3
Generously allowed region (%)	0
Disallowed region (%)	0


A
Relative Energy
RMS
250
200
150
100
50
0
-50
0
2
4
6
8
10
12
14
16
18
20


B


C

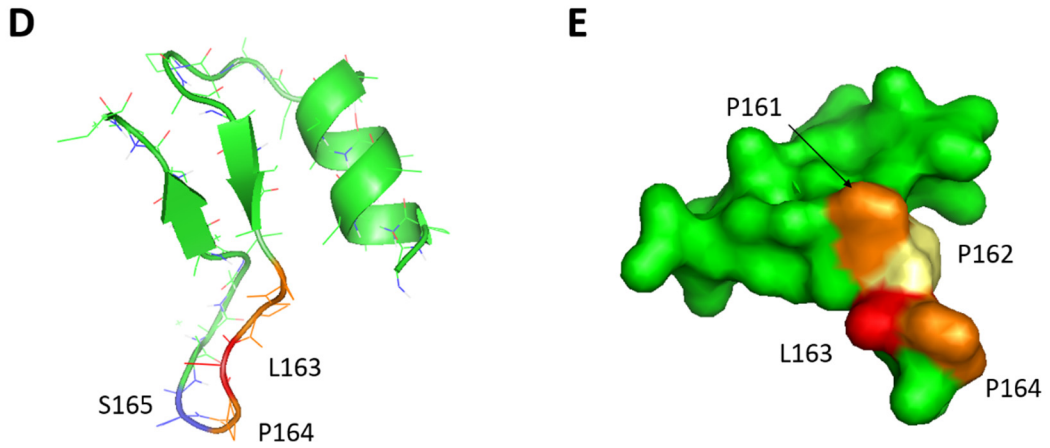


Figure 3.5: The solution structure of bound Vif SOCS-box (34 residues). (A) The RMS vs relative energy plot figure from CS-Rosetta calculation. (B) The alignment of ten models with the lowest energy after refinement. (C) The alignment to the α-helix of the crystal structure (PDB ID: 3DCG, chain E). The crystal structure is shown in violet. (D) The solution structure of Vif SOCS-box. Prolines on the PPLPS motif are coloured in orange. Leucine and serine are coloured in red and blue respectively. (E) The surface of the SOCS-box. The second proline in the PPLPS motif is coloured in yellow. The leucine is highlighted in red.

Our group has demonstrated that the PPLPS motif lying after the Lys157-Lys160 β-sheet binds to the C-terminus of EloB, and the peaks of Leu163, Ser165 and Val 166 fade away with the increasing ratio of EloBC (Bergeron, Huthoff *et al*. 2010). The solution structure reveals that the target short loop is composed of residues Pro164 and Ser165, and that residue Leu163 is sitting closer to the loop (figure 3.5D), suggesting the importance of these three residues. Specifically, these two β-sheets in the structure push Pro164 and Ser165 outwards so that they can easily interact with other surrounding molecules in solution to perform biological functions. A careful inspection of the surface of the SOCS-box structure can infer that the side-chains of the first and the third proline and the leucine are responsible for the interaction (figure 3.5E). This hypothesis was further demonstrated by biological assays in Chapter 6.

3.5 Summary

In this chapter the structure of SOCS-box bound state with EloBC is described. The main challenge in this study was the lack of chemical shifts and subsequently NOE assignment, which hindered the structure calculation by conventional methods. Various conditions were tested, and they failed to restore the broadened peaks. Due to this, CS-Rosetta provided an alternative method to generate structures. In theory, CS-Rosetta does a data-mining in fragment selection stage, by which a published structure might be referred to if the amino acid sequences are the same or homologous. In our case, the generated coordinates nicely agree with the crystal structure 3DCG, which demonstrates that the method provides reasonable results. More details about the interaction with EloBC are illustrated in Chapter 5.

CHAPTER 4: NMR STUDY OF ELONGIN BC COMPLEX

This chapter describes the solution structure of Elongin BC. The structure was calculated and reported by the same methods used in Chapter 3. Besides, NMR spectrometry was applied to study EloB mutants in order to identify the characteristics of EloB DVMK stretch at the C-terminal tail in the binding to SOCS-box.

4.1 ^{15}N-HSQC spectra of Elongin BC

EloB and EloC are an insoluble proteins if expressed in *E. coli* independently. Until now there have not been reports on the biophysical or biochemical studies of EloB or EloC individually. Thus the EloB-EloC complex has to be co-expressed and co-purified as described in Chapter 2. The ^{15}N-HSQC spectrum displayed in figure 4.1A reveals a partially-structured behaviour of the EloB-EloC dimer in solution. However, due to the large molecular weight (26 kDa) and exchange rates, peaks in the central region of the spectrum cannot be distinguished clearly, except those from flexible domains. Deuteration was used to replace a proportion of the non-exchangeable protons with deuterous. Briefly, in order to obtain deuterated sample, bacteria were grown in 100% D_2O instead of H_2O supplemented other normal chemicals for minimal media. Normal H_2O buffer was used during purification. Figure 4.1B confirms that the deuteration gives a much improved ^{15}N-HSQC spectrum compared to the regular one, showing good peak distribution and providing evidence for a good fold.

4.2 Elongin BC assignments

The original assignments of EloBC bound to the SOCS3 fragment (BMRB ID: 15606) were completed by Jeffrey Babon's group in 2008 (Babon, Sabo *et al*. 2008). Dr. Julien Bergeron finished the backbone assignment for EloBC bound to Vif SOCS-box. In order to

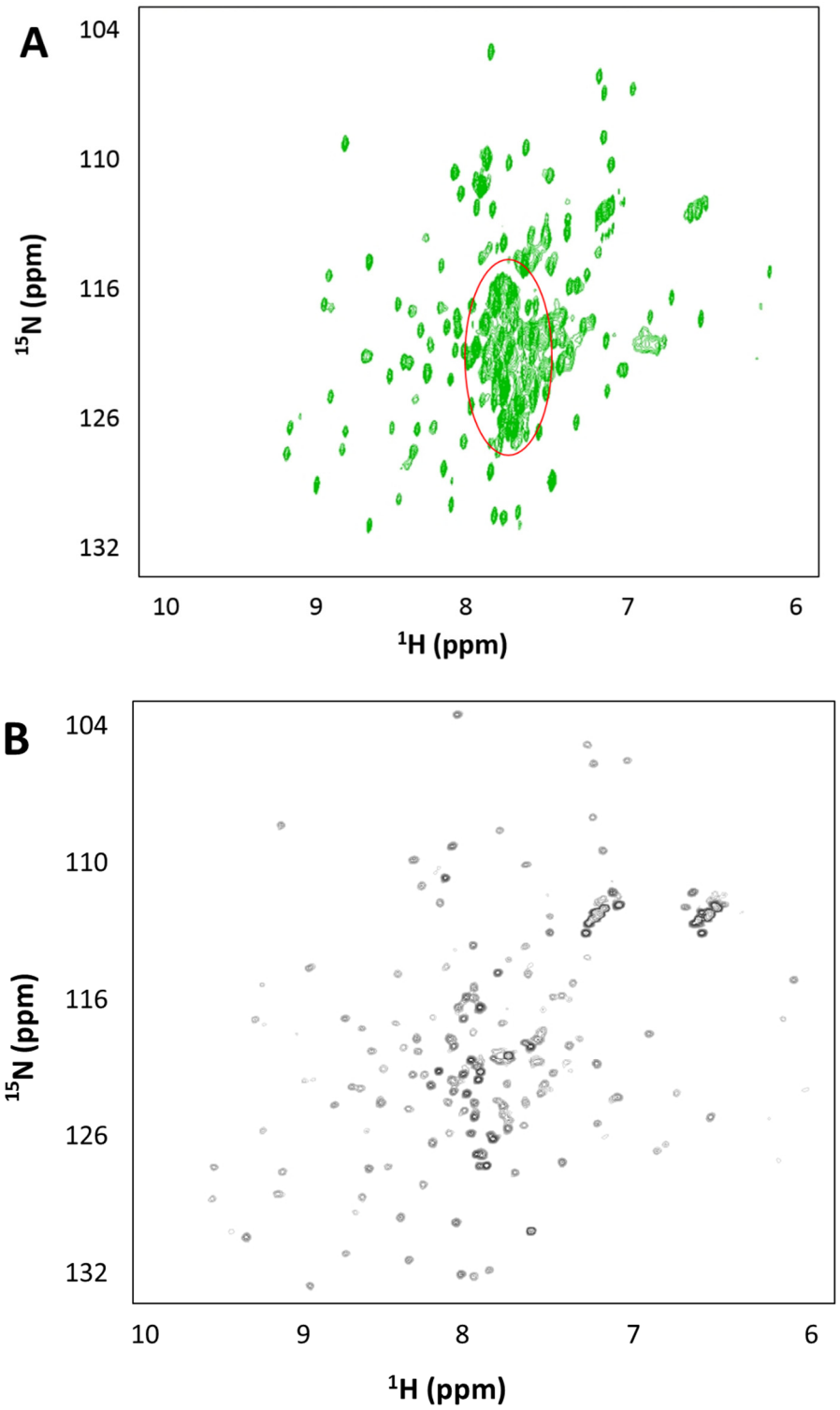


Figure 4.1: ^{1}H-^{15}N HSQC spectra of EloBC. (A) HSQC spectrum run on normal ^{15}N-labeled EloBC. Peaks in the red circle are from the EloB C terminus. Partial assignments of these peaks have been published (Bergeron, Huthoff et al. 2010). (B) HSQC spectrum on 100% deuterated ^{15}N-labeled EloBC.

obtain side-chain assignments, HCCH-COSY, HCCH-TOCSY and ^{13}C-NOESY-HSQC spectra were recorded. However, on account of the molecular weight of the complex (38 kDa.), 3D spectra recorded on regular EloBC sample in the complex gave only sharp peaks assigned to the mobile and unstructured C-terminus of EloB, residues 98 to 118 which could be assigned (figure 4.2). The absence of peaks for the less mobile regions meant that it was impossible to assign the side-chains for the rest of the dimer.

As for the EloC component, due to the same limitation affecting the assignment of EloB, the side-chains of EloC were not assigned. Considering that CS-Rosetta generates structures based on backbone assignment and that the EloBC structure has been solved, this disadvantage did not affect the *de novo* structural generation. Regardless of all the difficulties, the limited data still enabled one to identify the interaction mechanism between the PPLPS motif and EloB 101-104 DVMK stretch. The sequences of EloB and EloC are given in figure 4.4 with bioinformatic analysis.

In order to assign more EloBC resides and to confirm that there were no peaks corresponding to structural regions overlapped with other peaks, a selective labelling strategy was tried. 1-^{13}C-labeled amino acids were added into the M9 minimal media supplemented with ^{15}N-labeled ammonium chloride. Two different amino acids, proline and glutamine, were used due to the residual percentages in the sequences. A HNCO spectrum was recorded and the processed H-N plane of the spectrum is displayed in figure 4.3. This result reveals that there are no extra peaks appearing except for those from the carboxyl tail of EloB and a flexible element of EloC, proving that peaks corresponding to unassigned residues are absent from spectra due to line broadening.

4.3 Solution structure calculation of Elongin BC

4.3.1 Bioinformatic analysis

Similarly, we performed a hydrophobic analysis, a random coil index analysis and torsion angle prediction on EloB and EloC respectively to prepare the input file for CS-Rosetta

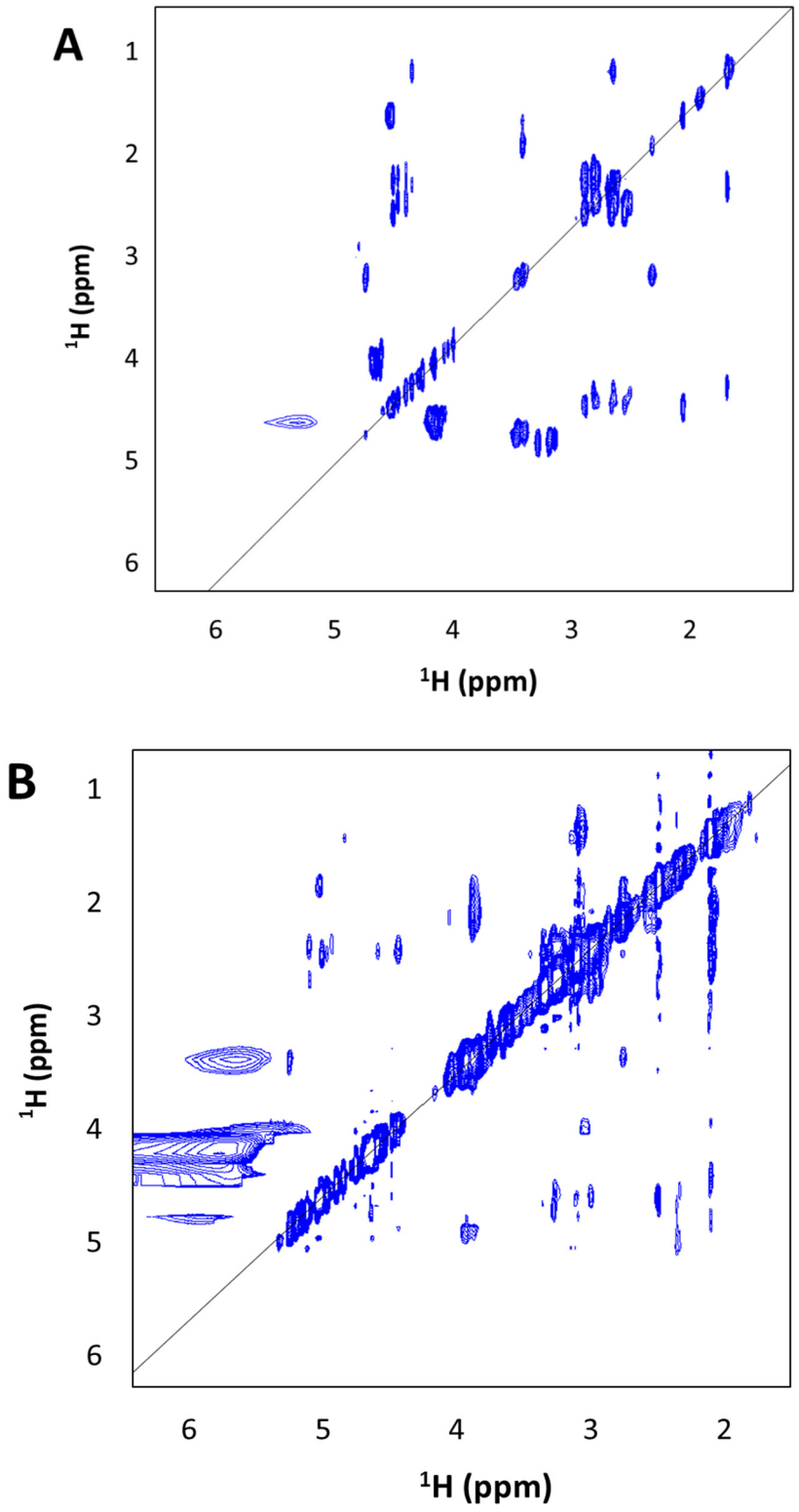
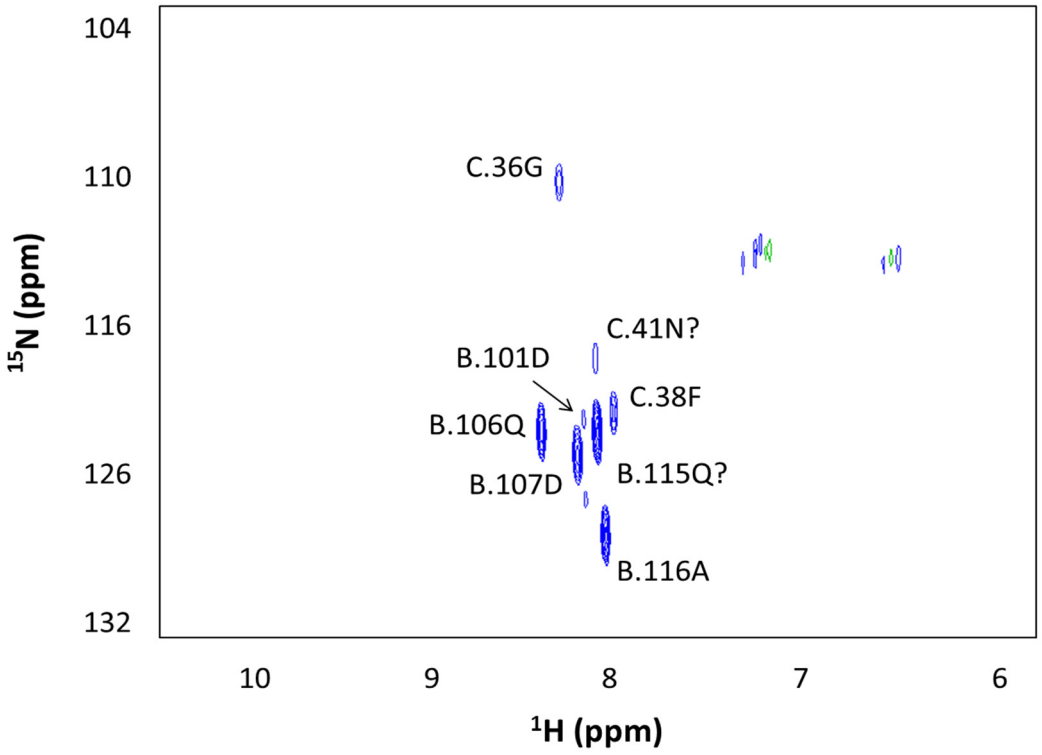


Figure 4.2: The ^{1}H-^{1}H projection of 3D spectra of EloBC in the complex. (A) HCCH-TOCSY. (B) ^{13}C-NOESY-HSQC.



4.3: The ^{1}H-^{15}N plane of HNCO collected for 1-^{13}C glutamine and proline-labeled EloBC. Assigned labeled residues are indicated. Due to the pulse-program, residues at the C-terminus to the ^{13}C-labeled amino acids can be observed on the ^{1}H-^{15}N plane. Peaks are assigned on the spectrum. B and C stand for chain B or chain C respectively. Question mark indicates unexpected peaks.

structure generation. The hydrophobicity analysis reveals that EloB and EloC contain many hydrophobic residues distributed throughout their sequences (figure 4.4A). EloC, especially, has a hydrophobic C-terminal tail, which is known to bind members of the SOCS-box family (figure 4.4C). The hydrophobic characters of the residues in these two proteins may make it difficult to express EloB or EloC independently in *E. coli*.

The RCI predication suggests that the Ubiquitin (Ub)-like-domain of EloB (residues 1-80) is structured and stable in solution (figure 4.4B). The predication on the C-terminal tail after residue 98, however, shows that the C-terminal tail is very disordered as demonstrated by the ^{15}N-^{1}H HSQC spectrum (figure 4.1A). Meanwhile the feedback from TALOS+ provides a good prediction for torsion angles except for the last 10 amino acids (figure 4.4E). These results indicate that the flexible C-terminal tail will not be considered during the alignment to the published structures after structure generation.

Although the assignments for EloC were not as complete as for EloB, both RCI predication and torsion angle prediction display that the full-length EloC is less disordered. The last two residues LD were removed because of the flexibility during the structural generation (figure 4.4D, F). The blank entries between each column are caused by the lack of chemical shifts. All chemical shifts were adjusted by RCI automatically for CS-Rosetta calculation in the next step.

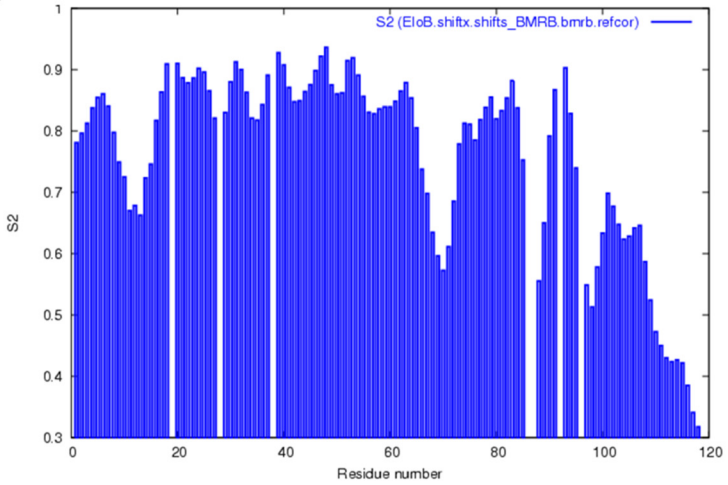
4.3.2 The structures of EloBC

As detailed above, the EloB structure contains a folded Ub-like domain whose second β-sheet (residue 11-16) forms an interaction surface to the EloC (figure 4.5A). Loop 68RPQA71 is another strand to hold the long α-helix of EloC (figure 4.5B). The modelling gives an unstructured flexible C-terminal tail as we expected. EloC, the component that is thought to have the only binding site of EloBC dimer to Vif SOCS-box, has three important domains. The second β-sheet and long α-helix 11-15 are involved in the binding to EloB. The C-terminus of EloC experiences a structural change to a hydrophobic α-helix that interacts with Vif BC-box (Stanley, Ehrlich *et al*. 2008, Bergeron, Huthoff *et al*. 2010). The structure of the dimer agrees very well with the published crystal structure, with a Cα r.m.s.d. of 1 Å over

A


MDVFLMIRRH KTTIFTDAKE SSTVFELKRI VEGILKRPPD EQRLYKDDQL LDDGKTLGEC
GFTSQTARPQ APATVGLAFR ADDTFEALCI EPFSSPPELP DVMKPQDSGS SANEQAVQ

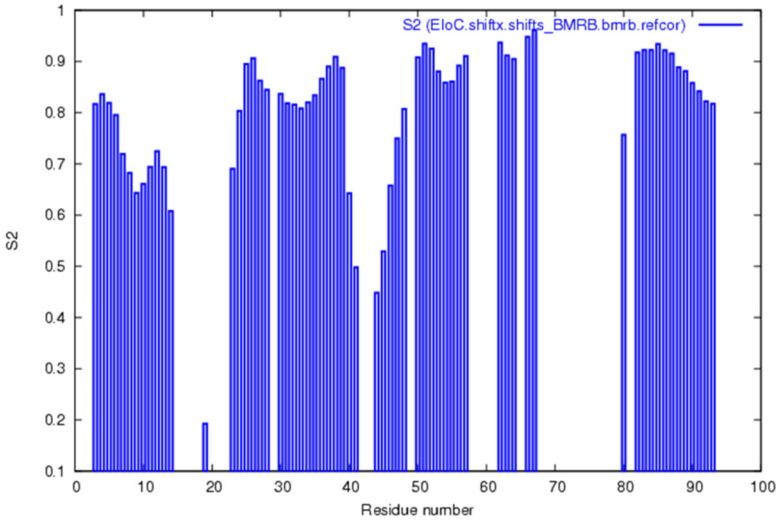
B


S2 (EloB.shiftx.shifts_BMRB.bmrb.refcor)
S2
Residue number
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
20
40
60
80
100
120

C


MYVKLISSDG HEFIVKREHA LTSGTIKAML SGPGQFAENE TNEVNFREIP
SHVLSKVCMY FTYKVRYTNS STEIPEFPIA PEIALELLMA ANFLD

D


S2 (EloC.shiftx.shifts_BMRB.bmrb.refcor)
S2
Residue number
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
70
80
90
100

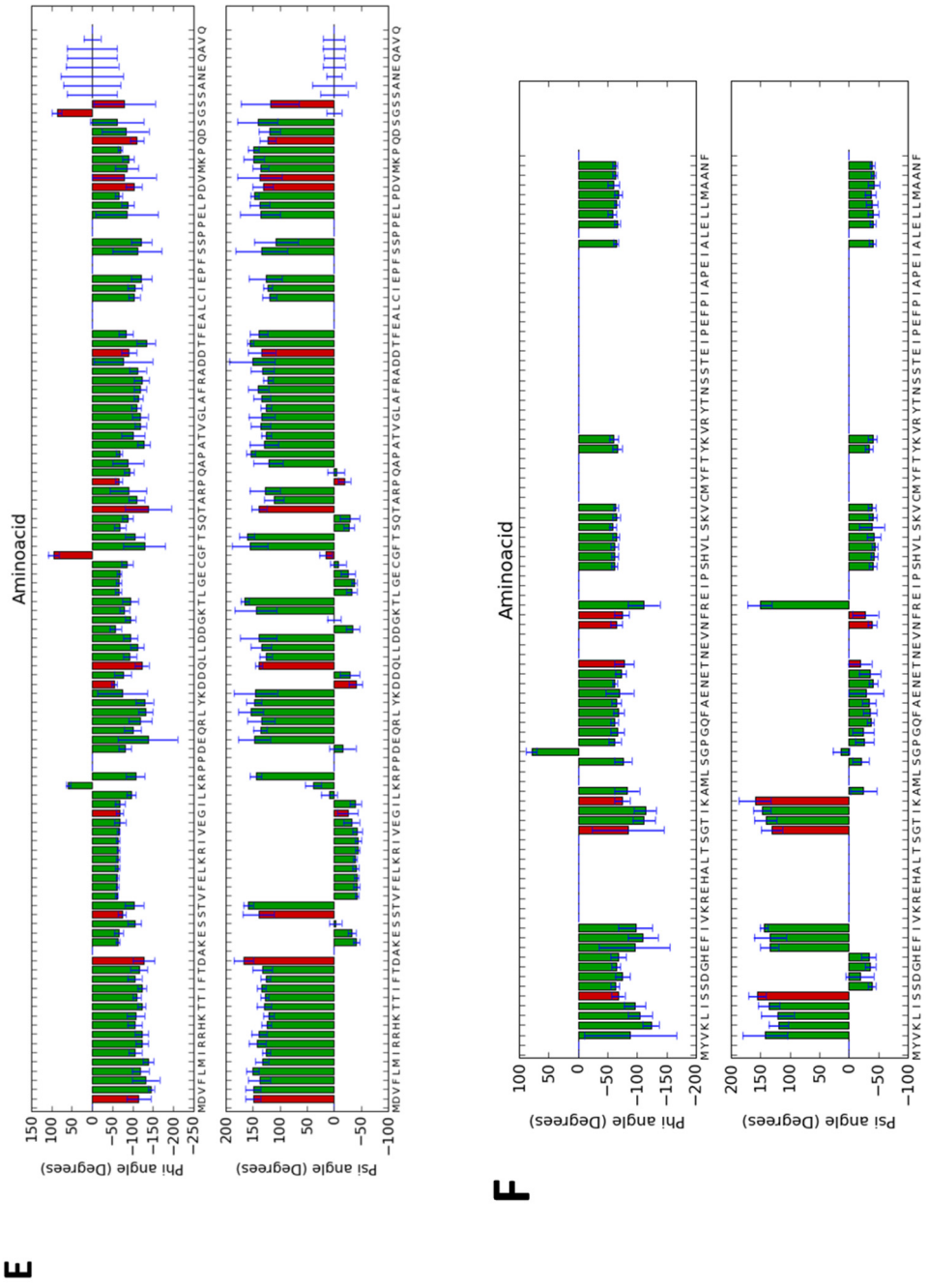


Figure 4.4: Amino acid sequence analysis of bound EloBC. (A) & (C) hydrophobic analysis on EloB and EloC respectively. Blue represents polar amino acids and green represents non-polar amino acids. The binding interface between EloB and EloC is indicated with a red line. (B) & (D) Flexibility prediction on EloB and EloC. (E) & (F) Torsion angles analysis on EloB and EloC. The last two residues of EloC were removed for CS-rosetta calculation.

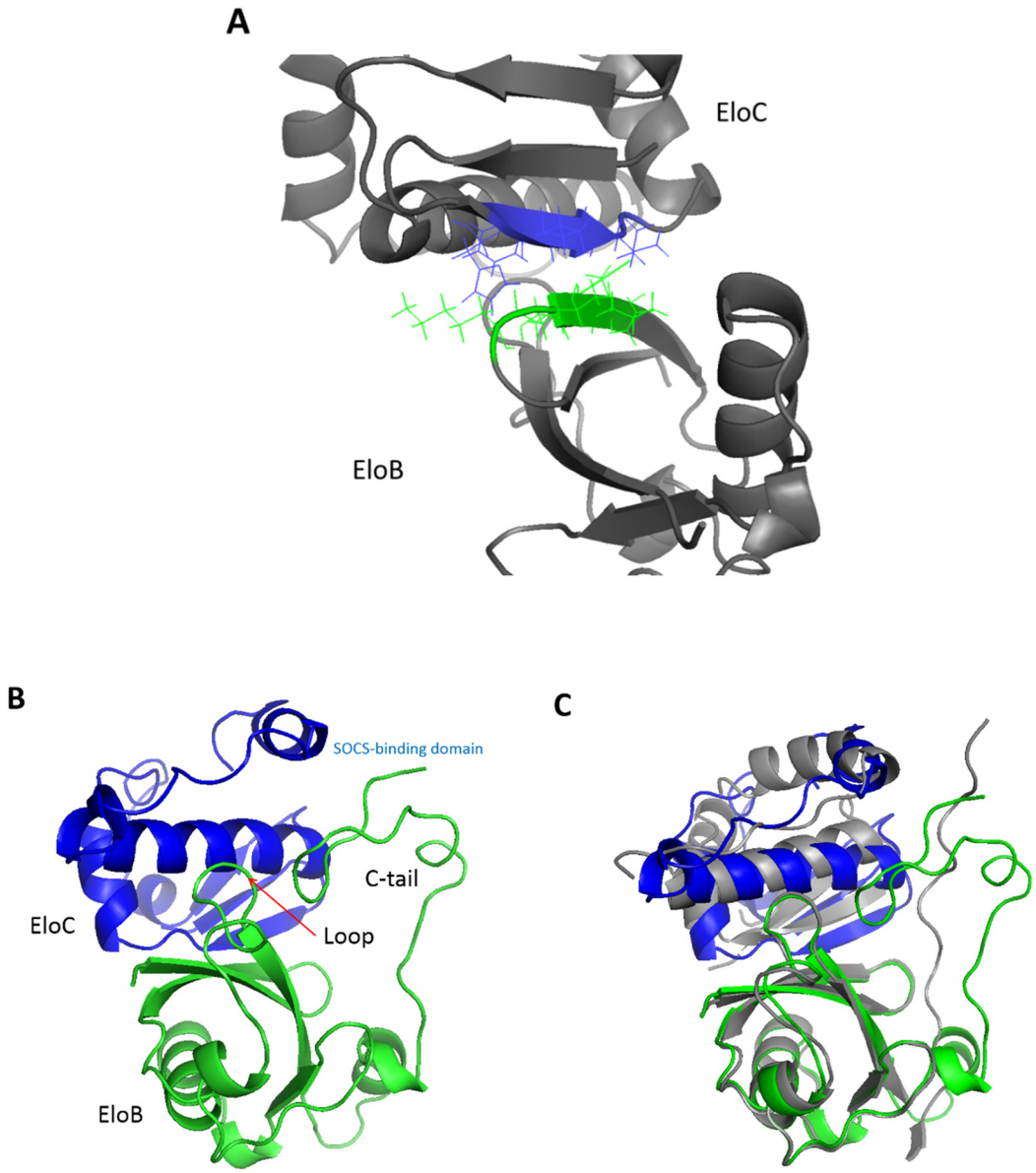


Figure 4.5: The solution structure of bound EloBC. (A) The interface between EloB and EloC. Residues 11-16 on EloB are coloured in green and residues 62-67 on EloC are coloured in blue. (B) The loop (residues 68-72) on EloB is another domain that binds EloC. (C) The alignment with the crystal structure (PDB ID: 3DCG, chain A & B). The grey structure are the EloBC crystal structure.

200 residues (figure 4.5C). The structure with the lowest energy was then used in the stage of docking the complex.

4.4 The character of EloB C-terminus in binding to Vif SOCS-box

It was found in our previous studies that residues 101 to 104 of EloB participate in SOCS-EloBC binding as well besides the C-terminus of EloC (Bergeron, Huthoff *et al*. 2010). We then enquired how these two domains interact and which residues may be critical for the binding. In order to observe the changes, we made four EloB mutants, each of which has a mutation site in the 101-104 stretch (figure 4.6A). ^{15}N-labeled sample was split into two sample fractions. Fraction I was mixed with unlabeled SOCS-box at 1:1 ratio and fraction II was diluted with NMR buffer to the same concentration. Briefly, if an alanine peak keeps changing during binding as well as the other three residual peaks, it suggests that this residual site does not determine the interaction, and that the interaction is not driven by side-chain interaction.

CcpNmr suite was used to detect peaks and to calculate peak heights. Some peaks that had been known not to be affected by the binding were chosen as references and normalized to the same height in two individual spectra. The final percentages were calculated from the height in the bound state relative to that in unbound state for each mutant.

Mutant EloB101 has a slightly different profile compared to WT. Only peaks from residues 101 and 102 have significantly decreasing intensities by 70% whereas residues 103 and 104 retain the same intensity (figure 4.6B). The ^{15}N-HSQC spectrum of this mutant displays two positions of those peaks from C-terminus which change when it is referred to the WT, confirming that this site mutation, to some extent, affects the rest of the sequence. Importantly, the decreasing intensity of 101A peak reveals that the interaction to the 101 site still exists. Interestingly, the other three mutants, provide the same profile of intensity changes as what we observed for WT (Bergeron, Huthoff *et al*. 2010). It shows that none of these residues are crucial, and that the side-chains do not play a role in this interaction.

A

```
EloB MGSSHHHHHH SQDPMDVFLM IRRHKTTIFT DAKESSTVFE LKRIVEGILK
     RPPDEQRLYK DDQLLDDGKT LGECGFTSQT ARPQAPATVG LAFRADDTFE
     ALCIEPFSSP PELPDVMKPQ DSGSSANEQA VQ

101A ---------------AVMK--------------VQ
102A ---------------DAMK--------------VQ
103A ---------------DVAK--------------VQ
104A ---------------DVMA--------------VQ
```

B

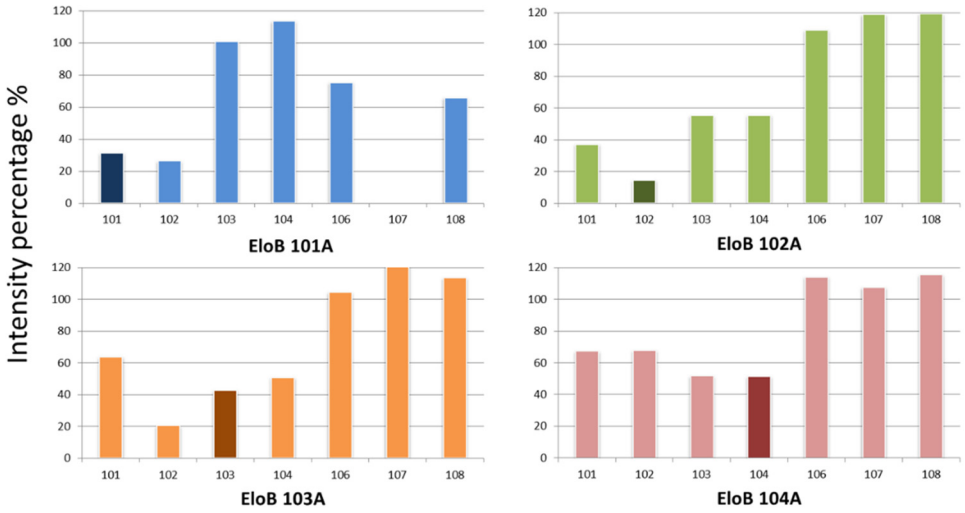


Figure 4.6: Peak analysis for the C-terminal element of EloB. (A) Schematic representation of the EloB construct (including the His-tag in the cyan box) used in this assay. The domain of interest is coloured in red and each mutation site is shown in blue. (B) Relative strength of the peaks assigned to residues 101 to 108. The percentage is given by the ratio of peak intensities in the spectra of the bound EloBC to those of the free EloBC.

4.5 Summary

In this Chapter I analyzed the characteristics of the EloBC amino acid sequences, assigned the side-chains of EloBC and calculated the structure of EloBC from our assignments. The structure generated from CS-Rosetta is in agreement with the crystal structure. This structure was used in later structural calculation of the SOCS-EloBC complex by HADDOCK. Results from NMR mutational study reveal that although the 101-104 stretch is critical for the EloB-SOCS binding, a single site mutation does not affect the binding. Moreover, it also implies that the interaction may not be related to any side-chain effect, because peak intensities from mutated alanine sites were decreased as well. The mechanisms under pinning these results are to be illustrated in Chapter 5.

CHAPTER 5: STUDIES ON THE SOCS-ELOBC COMPLEX

In previous two Chapters the structures of Vif SOCS-box and Elongin BC were presented respectively. This Chapter describes the entire structure of SOCS-Elongin BC complex and their interfaces. Meanwhile, biophysical methods were applied to quantify and characterize this interaction. The results suggested that the PPLPS motif did not bind to EloB carboxyl tail as tightly as we had expected.

5.1 Paramagnetic relaxation enhancement study

The binding between HIV-1 Vif SOCS-box domain and EloBC forms a stable heterotrimer with a 1:1:1 stoichiometric ratio determined by gel filtration analysis and SDS-PAGE (figure 5.1). In order to gain the distance information between these three components, NOESY spectra of the complex were recorded. However, due to several reasons mentioned in Chapters 3 and 4, a large number of inter-molecular NOEs were not observable using routine ^{15}N and ^{13}C-edited NOESY experiments, meaning that without sufficient distance restraints, the generated structure will not be accurate.

Considering that there are no cysteines in the SOCS-box peptide, and that the complex structure has a 1:1:1 ratio, it is an advantage to generate single-point mutations in this peptide so that the distances by PRE between the substituted amino acid in question and amino acids from other components of the complex, namely within EloBC, can be measured. Basically, in each molecule of the complex, a SOCS-box peptide was engineered with only one cysteine substitution, therefore the observed relaxation enhancement effect arose from the PRE molecule bound to the introduced cysteine. A paramagnetic relaxation enhancement procedure was applied to acquire more intermolecular restraints. Point mutations were placed at three widely separated sites in order to decrease the bias from having only a single mutation (figure 5.2A).

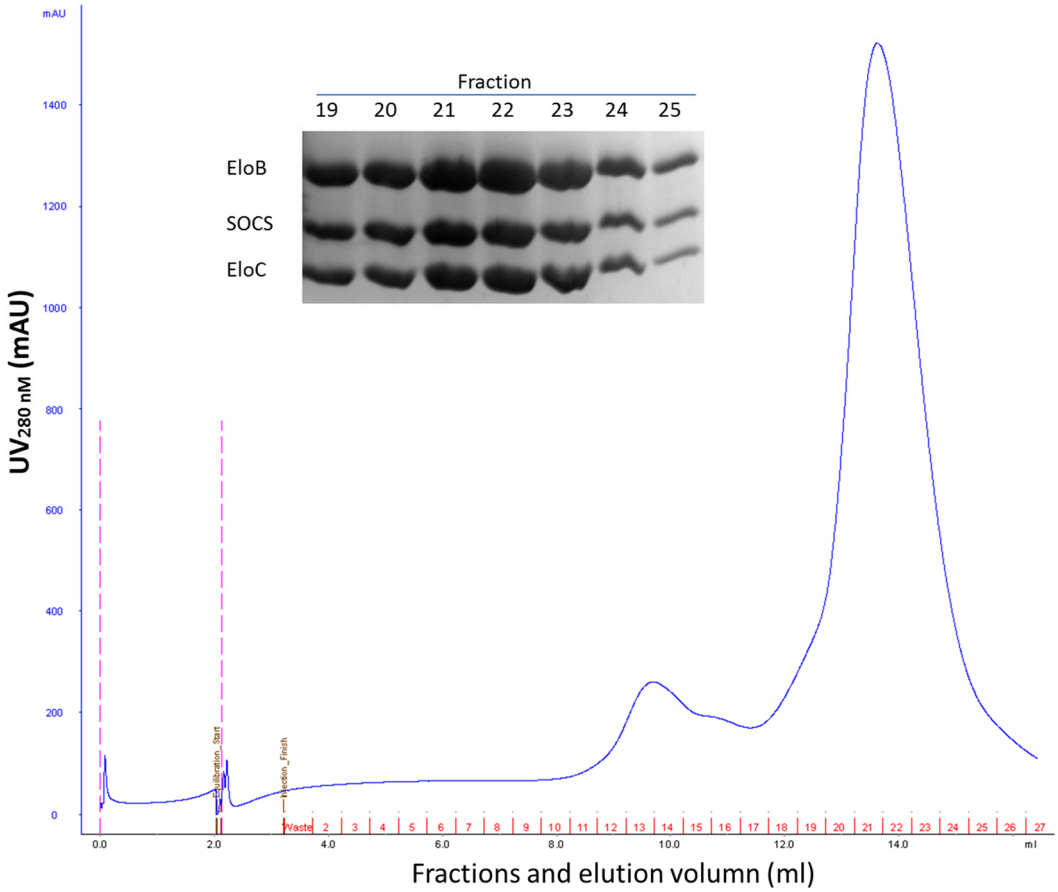


Figure 5.1: A purification profile of SOCS-EloBC complex. The SDS-PAGE gel (inset) shows the samples from the fractions 19-25. EloB, 16 kDa. SOCS-box, 11 kDa. EloC 10 kDa.

As was expected from other assays, peaks from residues 101DVMK104 were significantly weakened or invisible in the presence of each paramagnetic-labeled SOCS-box. This phenomenon was caused by both NMR perturbation effects and paramagnetic relaxation effects. Individually, the C-terminus of EloB was less impacted by the G143C mutant compared to the others (figure 5.2B). Paramagnetic-labeled Q158C mutant displays a clear clue to the possible positions of these EloB carboxyl residues in space (figure 5.2C). Moreover, the peak-intensity ratios from EloB carboxyl terminus reveal that the DVMK stretch was closer to the third labeled cysteine (R167C) of the SOCS-box (figure 5.2D). These intensity ratios were calculated and transformed into distance restraints by the method described in Chapter 2.

5.2 Overview of SOCS-Elongin BC Complex

The overlay of 20-solution-structures after applying the HADDOCK refinement protocol is displayed in figure 5.3A. The N-terminus of SOCS-box is indicated. EloB and EloC are coloured in blue and grey respectively. The structural statistics are collated in table 5.1. It is worth mentioning that the structural validation by CING (Common Interface for NMR structure Generation, for NMR data and structure validation) gave an over 75% reasonable structure score (ROG coding) (Doreleijers, Sousa da Silva *et al*. 2012), which was even higher than that of the validation on Jeffrey Babon's structures based on NOEs. Besides, PROCHECK reported that over 95% of residues were folded in the allowed regions. Considering these validation reports, since the solution structures are quite closer to the crystal structure, our structures are within the confidence limits regardless of the warning from CING.

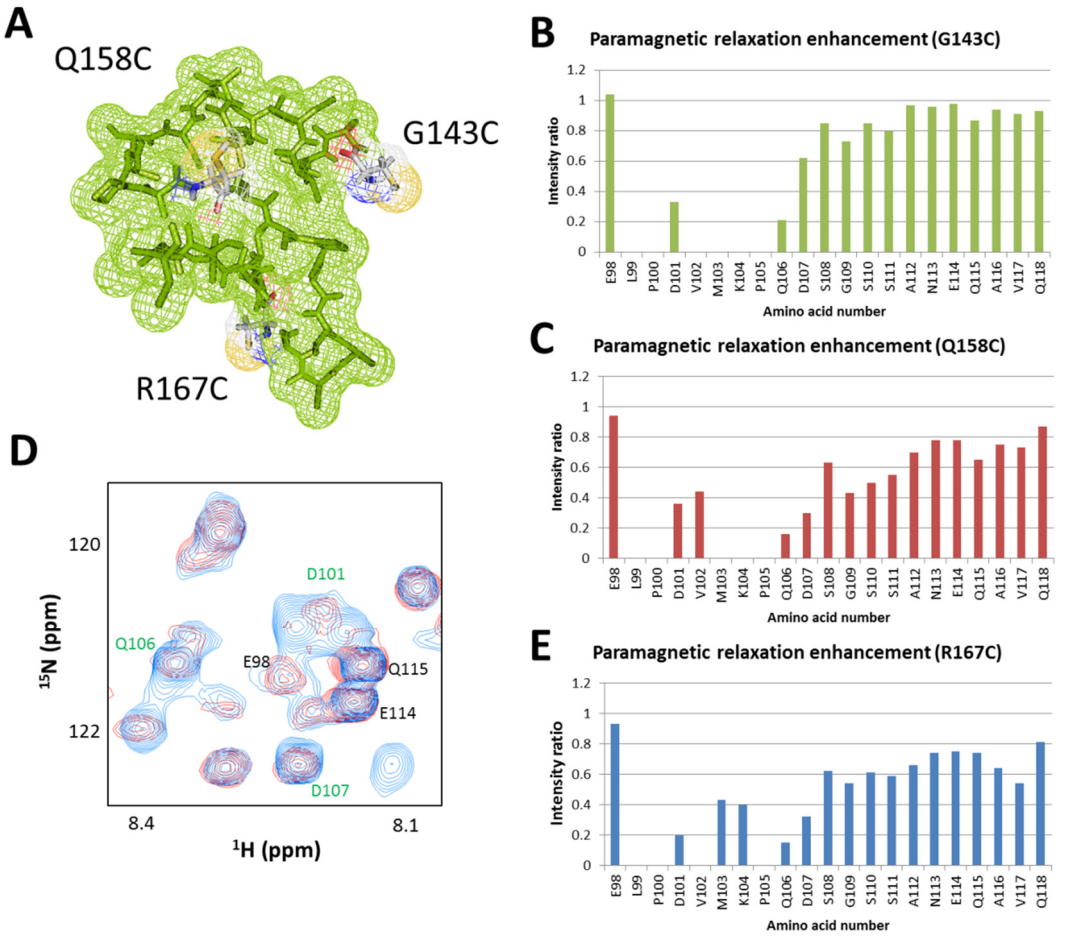


Figure 5.2: Long distance restraint measurements. (A) The three mutated sites in SOCS-box peptide which were distributed at widely separated positions in the peptide. G143C, Q158C and R167C are highlighted in white and the sulphur of the cysteine to be modified by label is coloured in yellow. Each predicted mutant structure is generated by Pymol in order to determine that the side chain of cysteine is not buried inside the molecule. (B) Intensity ratio of EloB C-terminus for G143C mutant. (C) Intensity ratio of EloB C-terminus for Q158C mutant. (D) A portion of the ^{1}H-^{15}N HSQC spectra of EloBC in the presence of the paramagnetic-molecule-labeled SOCS-box R167C peptide (red) or the diamagnetic-molecule-labeled R167C peptide (blue). Peaks assigned to the EloB C-terminal tail are indicated in green (significantly broadened) or in black (not affected). (E) Intensity ratio of EloB C-terminus for R167C mutant. The intensity of peaks from residues 101-104 are broadened due to the interaction regardless of the paramagnetic effect.

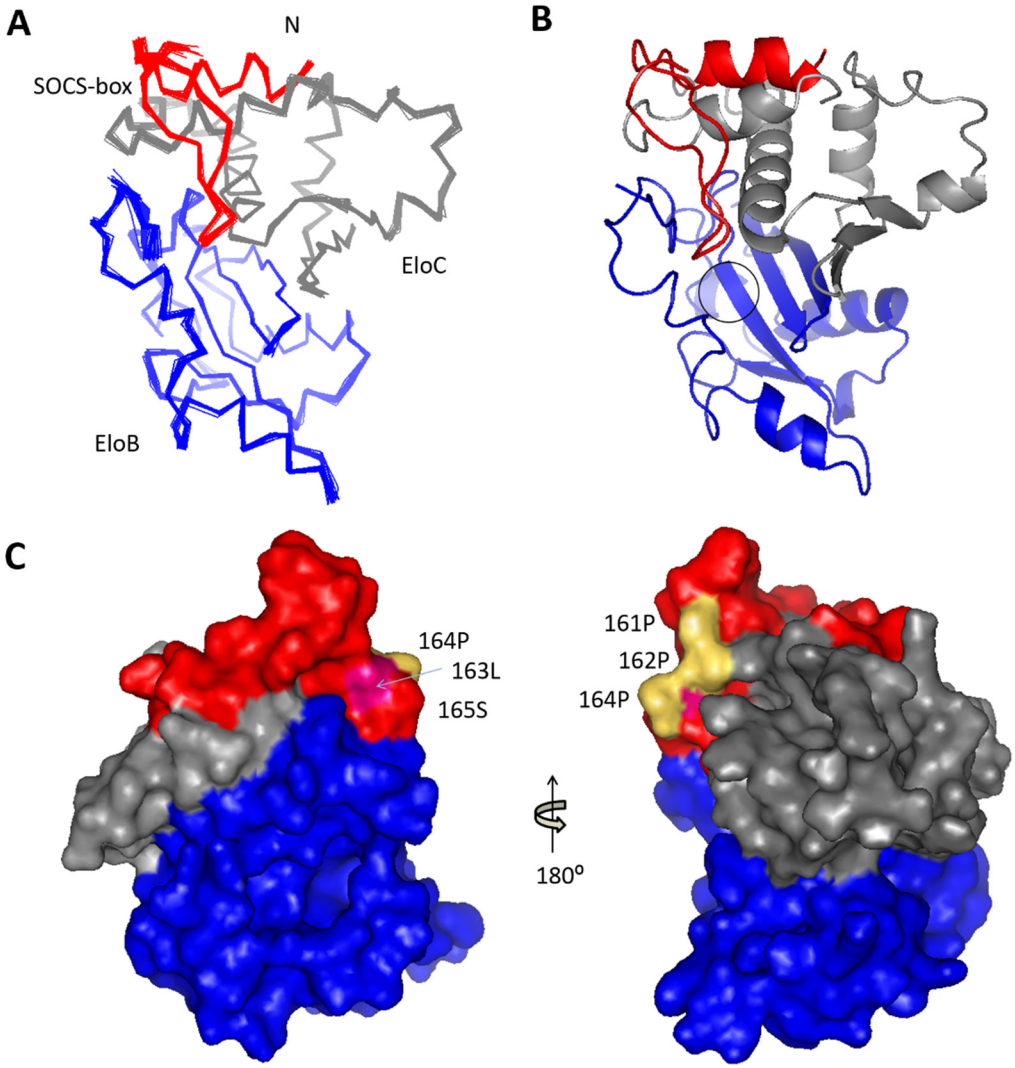


Figure 5.3: General view on the structure of the complex. (A) The alignment of twenty coordinates with the lowest energy after refinement. (B) The backbone cartoon of the final structure with the lowest energy. (C) The molecular surface of the complex. The proline-rich motif is displayed in yellow. EloB, blue. EloC, grey and SOCS-box, red.

Table 5.1. Structural statistics for the SOCS-EloBC complex

Average RMSDs from the mean structure	
Backbone average (Å)	0.26 ± 0.03
Heavy atom average (Å)	0.52 ± 0.04
HADDOCK statistics	
van der Waals energy (kcal mol^{-1})	-63.5 ± 2.4
Electrostatic energy (kcal mol^{-1})	-574.5 ± 15.9
Desolvation energy (kcal mol^{-1})	3.1 ± 3.3
Restraints violation energy (kcal mol^{-1})	0.3 ± 0.16
Buried Surface Area ($Å^2$)	2438.6 ± 46.0
Experimental PRE distance restraints	
Number	58
Distance violation (Å)	2.76 ± 1.90
Ramachandran plot	
Most favoured region (%)	82.3
Allowed region (%)	15.2
Generously allowed region (%)	1.4
Disallowed region (%)	1.1
Deviations from Ideal Geometry	
Angles (degree)	0.6
Bonds (Å)	0.003

The backbone cartoon of the structure with the lowest energy is presented in figure 5.3B. SOCS-box domain sits across the carboxyl tail of EloC. The remaining part of the SOCS-box is in a random coil structure. The small loop (163LPSV166) is close in space to the DVMK stretch of EloB, whose broadened peaks as observed in ^{15}N-HSQC spectra. Figure 5.3C presents the molecular surface of the complex. Three important residues on the PPLPS motif are highlighted in various colours. The side-chains of these three residues are clearly exposed on the interface co-formed by the PPLPS motif and the DVMK stretch of EloB, whereas the second proline is found buried inside the complex, implying that this residue is less important.

5.3 Binding sites between SOCS-box and Elongin BC

5.3.1 Interface between SOCS BC-box and the Elongin C carboxyl terminus

The structural model reveals that the SOCS-box crosses the EloC carboxyl α-helix and binds to the flexible EloB carboxyl terminus. Residues L145, A149, L150 of Vif SOCS-box and A99, L103 of EloC form hydrophobic interfaces respectively which drive the SOCS-EloC interactions (figure 5.4A, C), typical of protein-protein high-affinity binding (Schreiber 2002, Wolfe, Stanley *et al*. 2010). This tight hydrophobic interaction is the main driving force of SOCS-EloBC complex formation. This portion of the structure of the complex matches that of the crystal structure 3DCG very closely (0.41 Å).

5.3.2 Interface between proline-rich motif and Elongin B carboxyl terminus

Further investigations on the PPLPS motif shows that it is 164PS165 which interacts with the EloB DVMK stretch rather than the whole PPLPS motif (figure 5.4B). This ^{164}PS^{165} loop is stabilized by an anti-parallel β-sheet-like structure (shown in figure 3.5D). As for the EloB component, the C terminus was quite flexible in solution as shown by the line widths within the NMR spectrum (Bergeron, Huthoff *et al*. 2010), whereas in the presence of SOCS-box, the DVMK stretch experiences a conformational change and becomes partly structured by the

PPLPS motif (figure 5.4D). The distance measurement between these two domains suggests that the interface between 164PS165 and the DVMK stretch is formed by closer spatial positioning (around 5.5 Å) instead of bond connection or hydrophobic interaction. It indicates that this interaction is not as significant as observed in NMR perturbation experiments (Bergeron, Huthoff *et al*. 2010), and that the interaction observed by NMR may be driven by weak van der Waals forces.

5.4 Biophysical tests on the SOCS-Elongin BC Complex

The structure described above does not show any specific interaction mode between the PPLPS motif of the SOCS-box and the DVMK stretch of EloB. In order to confirm that there is no strong binding existing at the interface, ITC was applied to address this question and to quantify the interaction for different mutants.

Several constructs with single mutations in the PPLPS motif were made. The reason for replacing proline with a serine rather than an alanine is that serine, which is a hydrophilic acid with a small side-chain group, helped improve the solubility of SOCS-box (figure 5.5A). Meanwhile, this substitution also changed the character of the molecular surface to a hydrophilic surface, which provided more significant difference in this biophysical study. Each sample for ITC was purified through a gel filtration column to remove self-associated polymers of SOCS-box peptide. The concentration of SOCS-peptide was determined by Nanodrop spectrometer whereas that of EloBC was determined by BioRad due to its low extinction coefficient at 280 nm.

Figure 5.5B shows that the thermal changes could be measured calorimetrically by the binding reaction upon the titration of the SOCS-box with EloBC. Titration points were fitted to Daniel Koshland's theory of enzyme-substrate binding model to obtain thermodynamic parameters. The plotted single-binding curve appears that the middle point (half-maximum binding) of each curve is around 1.0, confirming that protein A, namely SOCS-box or its mutants, and protein B, namely EloBC, interacts at 1:1 ratio, agreeing with the profile from

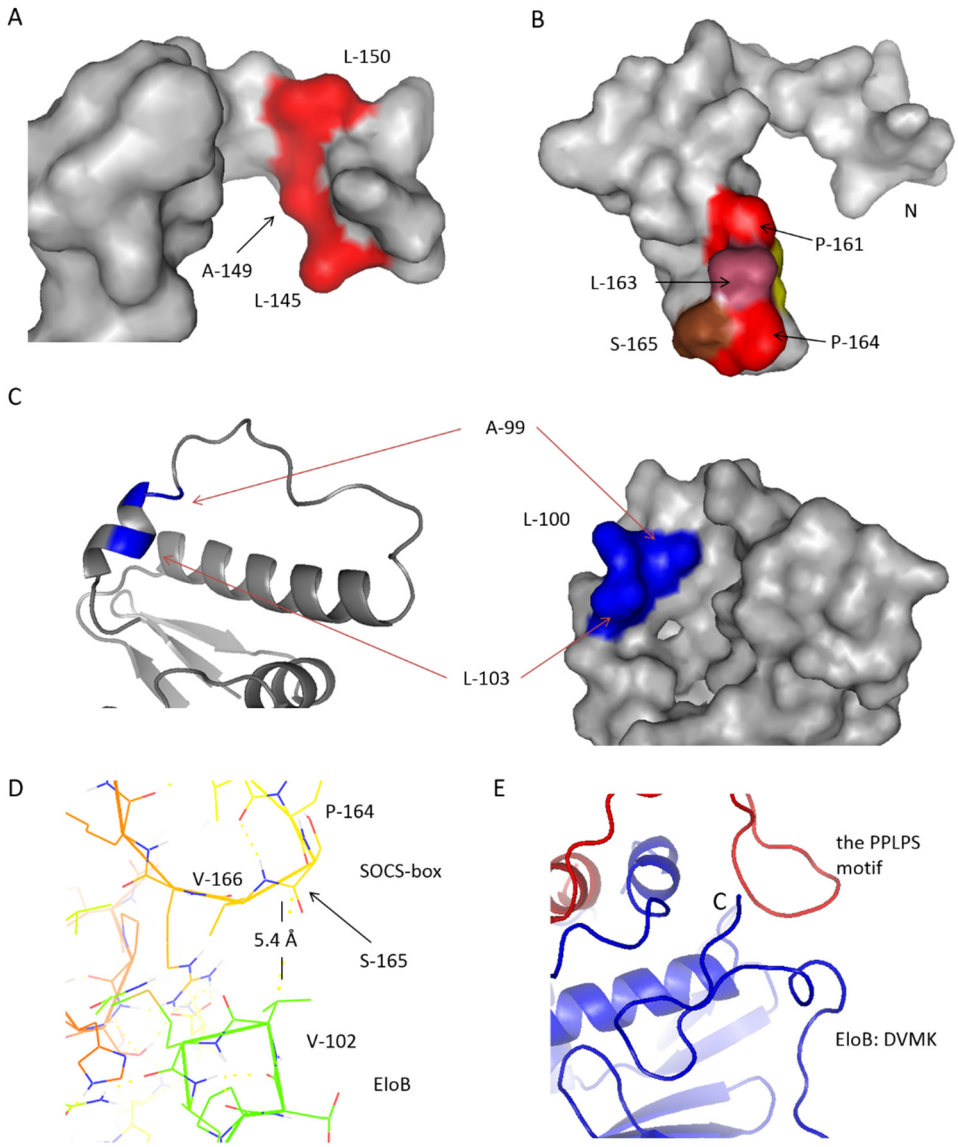


Figure 5.4: Interaction between the SOCS-box and EloBC. (A) The binding surface of the BC-box of the SOCS-box. L145, A149 and L150 form a hydrophobic surface which interacts with EloC C-terminus. (B) The proline-rich motif downstream from the SOCS-box is coloured. Residues are highlighted in various colours. The second proline (residue 162) in yellow is buried in the complex. (C) The binding surface of EloC C-terminal α-helix. Resides on the interface are coloured in blue. (D) The interface of EloB DVMK stretch and the Vif proline-rich motif. The distance between two motifs is presented. (E) The C-terminus of EloB, including the DVMK stretch. Blue, EloB. Red, SOCS-box.

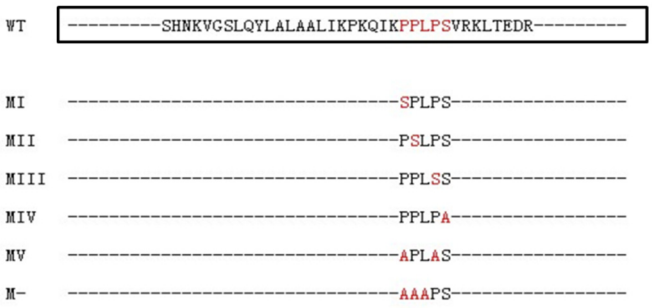
the gel filtration column (figure 5.5A).

The dissociation constant describes the binding affinity between molecules. The higher its value, the lower the binding affinity. Interestingly, the dissociation constants are not changed significantly by a single mutation with a maximum difference of 0.5 μM (table 5.1). This difference is usually regarded that the values of the binding constants of all target molecules are at the same level. In this case it means that a change on a single point does not affect the binding.

The ITC data also reveal that calculated $-T\Delta S$ derived from free energy and enthalpy differs among mutants (table 5.1). However, to be critical, it cannot be simply concluded that residue substitution influences the interaction for two reasons:

1) Hydration or dehydration property is another factor that should be addressed carefully in the conclusion on ITC data. This factor is usually used to describe a hydrophobic or hydrophilic force. Also it is used to explain interactions in charged conditions (Bjelic and Jelesarov 2008). This may be the case for the proline-serine replacement.
2) Generally, ITC is a method that sensitively detects the thermodynamic change of a whole target system, namely the SOCS-EloBC binding system in this work. It detects macroscopic changes rather than microscopic details on a system. It was found in the recent studies that slight modification on a site away from an active site can increase or decrease the binding affinity (Tzeng and Kalodimos 2012). This long-distance effect is thought to be led by conformational entropy. Similarly, the thermodynamic change observed in this case may be caused by the difference of internal conformational entropy rather than the direct binding energy. That could be why the dissociation constant remained the same whereas $-T\Delta S$ changed.

A



B

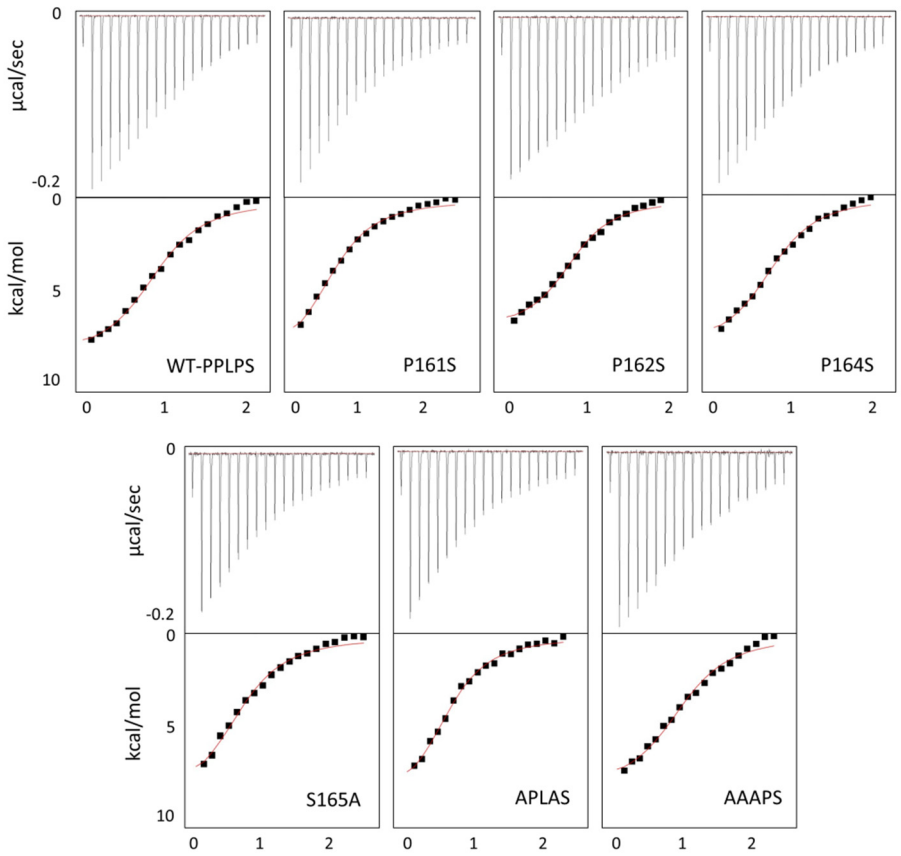


Figure 5.5: (A) Schematic representation of the SOCS-box constructs used in this assay. Mutated residues are coloured in red. (B) The data of ITC binding assay. Top, plotted curves according to the single-bind model. Bottom, raw data.

Table 5.2, Thermodynamic characterization of SOCS-EloBC interaction

	K_d (μM)	K_a (μM^{-1})	ΔH_{obe} (kcal/mol)	ΔG (kcal/mol)	$-T\Delta S$ (kcal/mol)
WT	1.23	0.82 ± 0.06	-10.2 ± 1.6	-7.9	2.3
SPLPS	2.14	0.52 ± 0.18	-8.3 ± 2.3	-7.6	0.7
PSLPS	1.24	0.86 ± 0.24	-8.8 ± 2.0	-7.9	0.9
PPLSS	1.36	0.77 ± 0.19	-10.7 ± 2.8	-7.8	2.9
PPLPA	1.18	0.89 ± 0.24	-8.6 ± 0.3	-7.9	0.7
APLAS	1.10	0.97 ± 0.28	-8.8 ± 0.2	-8.0	0.8
AAAPS	1.50	0.68 ± 0.11	-9.8 ± 1.7	-7.9	1.9

5.5 Summary

Here the first solved structure of the Vif SOCS-EloBC complex including the proline-rich motif on Vif component has been presented. SOCS-box primarily binds to EloC by a hydrophobic interaction in the BC-box and EloC binding surface. The proline-rich motif on SOCS mediates another weak interaction with the EloB DVMK (residues 101-104) stretch, suggesting that it is the α-helix of EloC that first drives the induced-folding binding of Vif, followed by the interaction between the PPLPS motif and EloB. The side-chains of the first and third prolines (residues 161 and 164) are exposed in the solution whereas the side-chain of the second proline is buried in the protein. The structure of the complex provides a detailed insight into the function of the Vif proline-rich motif. In the next stage biological assays were used to verify the structural model and to identify the function of each residue in cells, which will be discussed in the next Chapter.

CHAPTER 6: FUNCTIONAL STUDIES OF THE VIF PROLINE-RICH MOTIF

In previous chapters biophysical, biochemical and structural biological methods were applied to investigate the mechanisms behind the Vif SOCS-Elongin BC interaction. In particular, the solution structure of the complex was revealed. In this Chapter, biological assays were used to further identify the details of this interaction.

6.1 The role of each residue in the PPLPS motif

6.1.1 Expression of APOBEC3G and Vif plasmids in mammalian cells

HIV-1 vectors have been proven useful tools to study the early events in HIV-1 infected cells (Lee, Malim *et al*. 2008). The assay consists of a VSV-G pseudotyped HIV-1 based GFP reporter vector that expresses GFP upon completing the early replication steps in infected cells in a single round. Using flow cytometry it is possible to determine the infectious titer of the vector in the presence or absence of A3G. With help of this system it is possible to compare the anti-A3G activities of wild type Vif or Vif mutants that are produced along with the vector in presence of A3G.

In this system, pseudovirus that packages A3G protein along with a RNA sequence containing GFP cDNA and a packaging signal sequence were generated from 293T cells, and used to challenge 293T cells instead of TZM cells. In infectious cells A3G from the virus binds onto the promoter of GFP mRNA, thus blocking the expression of GFP in cells, from which the infectivity of a virus can be deduced, i.e. the more intensive the fluorescent signals are, the higher the infectivity. This method is therefore different from the method our group used before (Bergeron, Huthoff *et al*. 2010),

Considering the amount of DNA for the transforming efficiency and the toxicity of PEI to cells, we first asked how much A3G was required to significantly suppress the infectivity of HIV-1 GFP vector in this assay. The amount of HA-tagged A3G was co-transfected with a

gradient of the pseudovirus/GFP-expression plasmids in 293T cells. Virus-containing supernatant was collected and used to infect 293T cells. pHA-CMV, the parental vector of pHA-A3G, was used as negative control. As shown in figure 6.1A & B, increasing the amount of A3G plasmids to 0.5 μg could significantly suppress the infectivity to approximately 3%, corresponding to around 50-fold drop in infectious titer as compared to the negative control. Titration of A3G in this system caused only a weak decrease in virus production at high A3G concentration as judged by western blotting against p24 in the viral supernatant, suggesting that the observed antiviral effect predominantly is exerted by the activity of A3G in newly infected target cells, as described before (Lee, Malim *et al*. 2008). Similarly, by the titration with WT Vif plasmids, approximately 0.12 μg of NL4.3 Vif plasmids was enough to counteract the function of A3G (figure 6.1 C & D).

In order to study the function of each residue in the PPLPS motif later, a series of Vif mutants were constructed by a Stratagene Quick Change strategy. Vif plasmids were extracted using the QIAGEN Maxiprep kits. At the same DNA concentration read from Nanodrop spectrophotometer, DNA agarose gel confirms that the total amount of DNA for each mutant aliquot was the pure plasmid of interest without any observable DNA contamination (figure 6.1C).

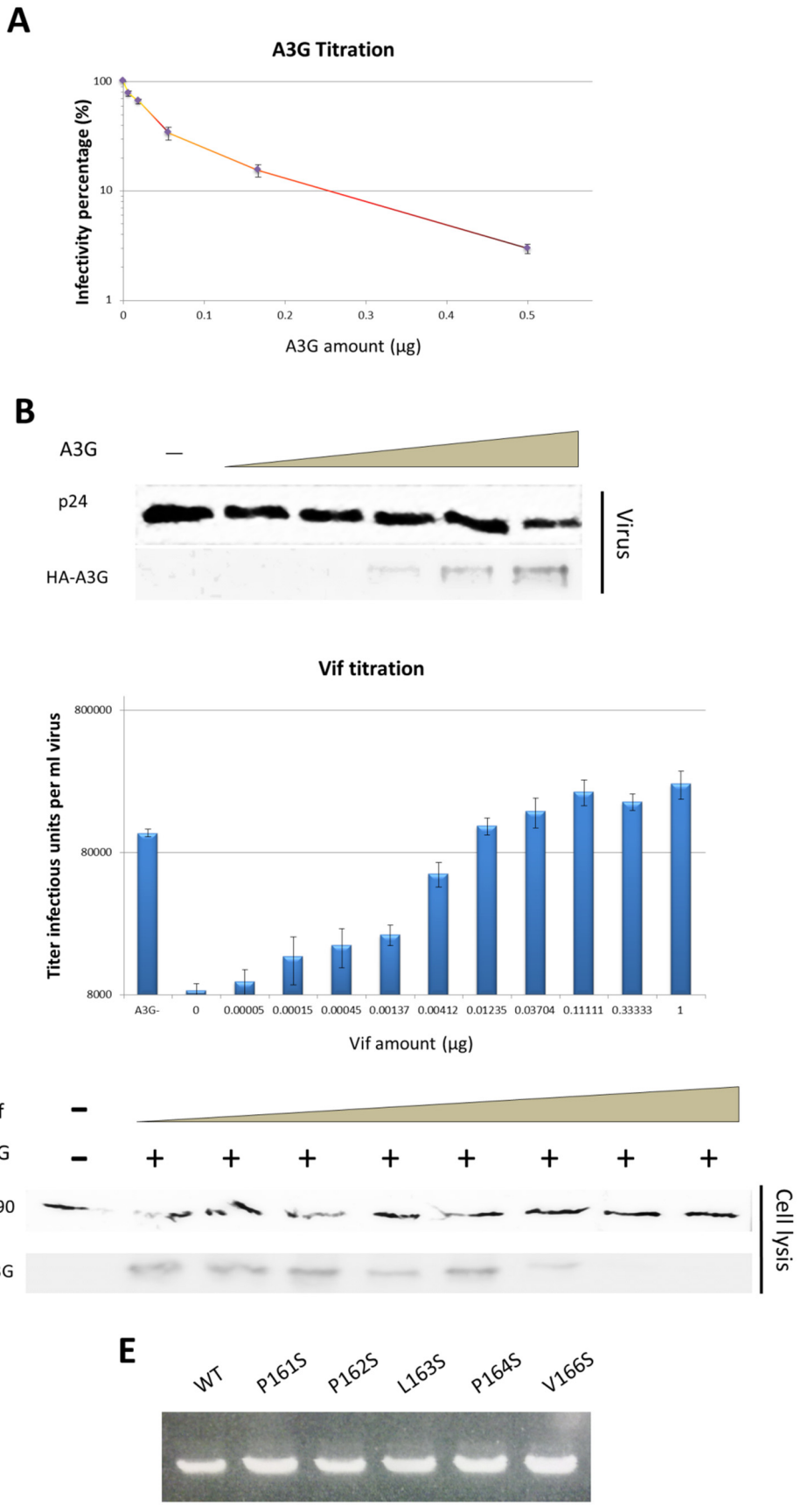

A
A3G Titration
Infectivity percentage (%)
100
10
1
0
0.1
0.2
0.3
0.4
0.5
A3G amount (µg)
B
A3G
—
p24
HA-A3G
Virus
C
Vif titration
Titer infectious units per ml virus
800000
80000
8000
A3G-
0
0.00005
0.00015
0.00045
0.00137
0.00412
0.01235
0.03704
0.11111
0.33333
1
Vif amount (µg)
D
Vif
A3G
Hsp90
HA-A3G
Cell lysis
E
WT
P161S
P162S
L163S
P164S
V166S

Figure 6.1: Expression of APOBEC3G in mammalian cells. (A) Infectivity of ΔVif HIV-1 pseudovirus with a titration of A3G. (B) Western blot of the virus supernatant used to challenge 293T cells. Gag p24 was used as a control of viral amount and as a loading control. (C) Infectivity of HIV-1 pseudovirus with a titration of Vif WT. (D) Western blot of the virus supernatant used to challenge 293T cells. Hsp90 was used as a control of cell amount and as a loading control. The second sample from the right is corresponding to the titration of 0.11 μl in C. (E) According to the reading from the Nanodrop spectrophotometer, same amounts of DNA were loaded onto the agarose gel. The same amount and purity of target plasmids are displayed

6.1.2 Effects of Vif mutations in A3G restriction

The function of different Vif mutants were then assessed by generating VSV-G pseudotyped HIV-1 GFP reporter vectors in 293T cells co-transfected with a plasmid encoding for A3G or control and a plasmid encoding NL4.3 Vif or Vif mutants. In one data set each condition was tested in triplicate. Three data sets were collected independently and a typical profile was presented.

The results, displayed in figure 6.2A, reveal that the first proline on the PPLPS motif and the valine downstream the motif are less important than other residues in the motif for Vif function. Changing the second proline to serine (P162S) does not decrease viral infectivity, which is in agreement with the structural information that the buried side-chain of this proline actually does not participate in functional interactions. Interestingly, the L163S and P164S mutants considerably lost their ability to rescue viral infectivity, suggesting a critical role in Vif function. However, the western blot on virus-producing cells shows that compared with WT Vif, L163S mutant had a slightly lower expression level compared to NL4.3 wild type Vif whereas P164S mutant was close to the limit of detection by immunoblotting (figure 6.2B). Now that original plasmids were checked before biological assays (figure 6.1C), this phenomenon may be caused by two reasons: 1) The transfected P164S plasmid has a very low expression yield in cells. 2) Vif mutant P164S is very unstable, and thus digested by proteases. Therefore, it cannot be simply concluded that P164 is critical for viral infectivity regardless of the fact that P164 is essential for Vif stability. Interestingly, L163S mutant Vif expression levels were only slightly reduced compared with wild type Vif (figure 6.2B), yet its anti-A3G activity was reduced by ~10-fold, suggesting that L163 is a critical residue for anti-A3G function.

Residue L163 was of interest given that substitution with serine caused a substantial defect in anti-A3G activity but not in Vif expression. To investigate the conservation of this residue in natural HIV-1 isolates we interrogated the Los Alamos HIV-1 sequence database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html). The majority of sequences have preserved a leucine at position 163, however, there were two isolate, B.JP.1995.PT6_i.AB034473 and B.US.1996.USPI90770EI197y96011pcWG2B4.JN024500, which had S163, suggesting that there may be variations in anti-A3G activities of natural

occurring Vif variants in circulating HIV-1 strains. The alignment of Vif sequence is attached in the appendix III (presented in an electronic version). We therefore tested a variant that appeared to a higher percentage in the database, F163. From a biochemical viewpoint, phenylalanine has a side-chain whose hydrophobicity and size are similar to those of leucine (although phenylalanine is aromatic). The same biological assays were applied again. It shows that even though the amount of transfected plasmid of L163S mutant was raised five fold to a level where the detected expression yield was similar to that of WT, viral infectivity still declined because of the presence of A3G (figure 6.3A & B). The L163F mutant, however, preserved Vif function and rescued the virus. These results confirm that the side-chain of residue 163 is critical for Vif function, and the loss of the side-chain prevents Vif from counteracting A3G.

As for the P164S, a titration on the mutant plasmids was performed. Figure 6.3C & D reveals that an increasing amount of the mutant did not affect the viral infectivity. Although from group 4 (10*P164S) it might be argued that this proline is important for Vif function on neutralization of A3G, it is not very reasonable or meaningful to draw such a conclusion because this Vif mutant was very unstable in cells, which means in nature it seldom exists. Interrogation of the Los Alamos HIV-1 sequence database revealed that P164 is very conserved, however, there are some sequences (B.CA.1998.CANB4FULL.AY779554, B.CA.1998.CANB5FULL.AY779555) with S164, again suggesting that natural circulating HIV-1 strains may have variations in their anti-A3G activity or stability of Vif. One would predict that HIV-1 strains with Vif S164 would express an unstable Vif, however, secondary compensating mutations rescuing Vif stability may complicate interpretations.

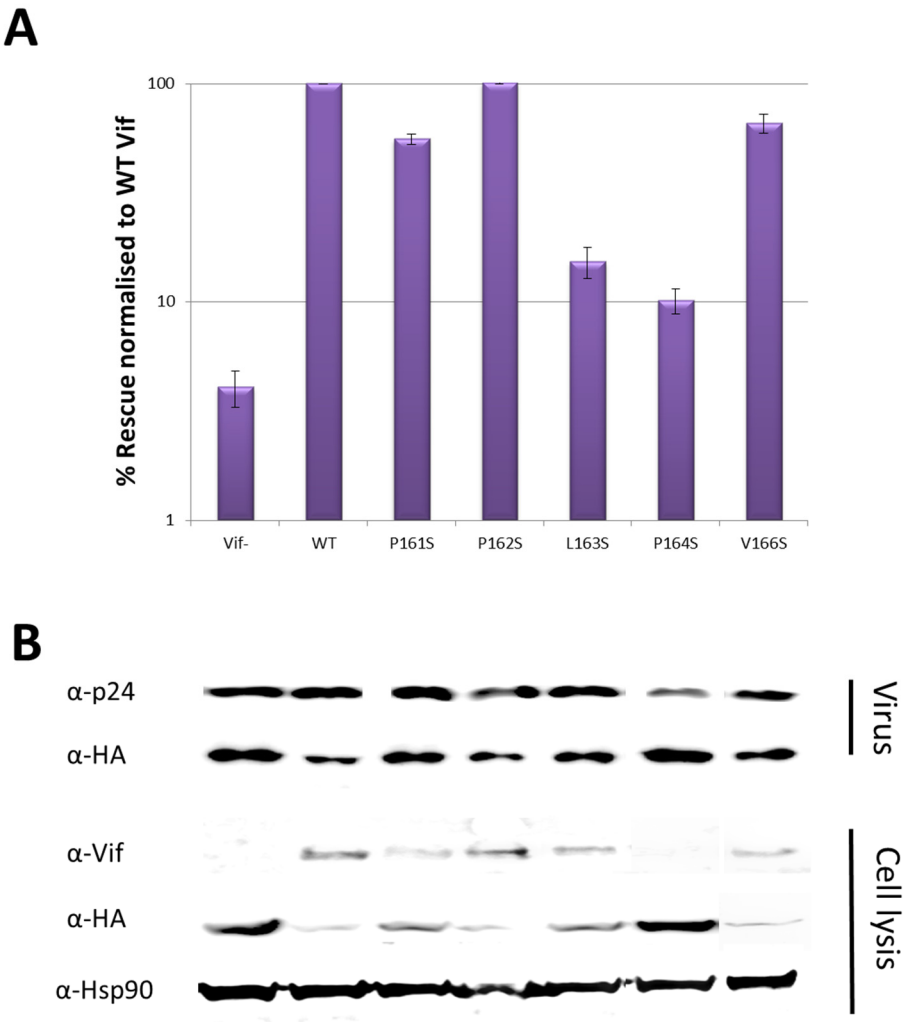


Figure 6.2: 163Leu and 164Pro are important for Vif function. (A) The function of Vif and its mutants was measured by the capacity to rescue virus in the presence of the anti-virus protein, A3G, in a single-cycle infectivity assay. The data are represented as relative infectivity normalized to WT Vif. The mean of three measurements was used, with error bars showing the standard deviation. (B) Immunoblot the virus supernatant and whole cell lysates from the producer cells. Gag p24 from virus was used for the control of viral amount. Hsp90 is shown as a loading control.

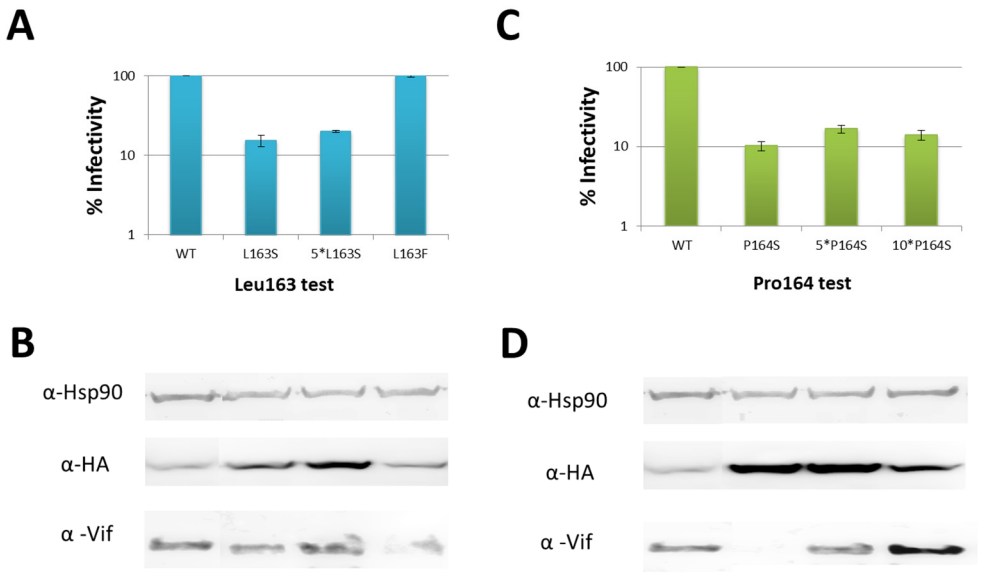


Figure 6.3 Single residue studies. (A) L163 is crucial for Vif function. The infectivity of HIV-1 pseudovirus with different amounts of Vif L163S plasmids or L163F plasmids. Plasmids pcDNA3.1 was used to keep the total amount of DNA same during transfaction. (B) Immunoblot whole cell lysates from the 293T producer cell lysis. (C) The plausible role of the third proline in the PPLPS motif. The infectivity of HIV-1 pseudovirus against the increasing amount of P164S mutants. (D) Immunoblot 293T producer cell lysis.

6.2 Roles of each residue in the PPLPS motif in forming the ubiquitination complex

The previous results suggested that L163 and P164 are more important than the first and second prolines. Now we were interested in whether these mutations would impact on the binding to the ubiquitination complex. Thus, co-IP of Cul5 with Vif and its mutants was used to study the mechanism and structure-function correlations.

The results (figure 6.4) firstly show that the P164S mutant experienced a low expression in this assay as well as in the infectivity assay, demonstrating its instability. Interestingly, the three single-mutation proteins, P161S, P162S and L163S, retained the ability to recruit the ubiquitination complex as did NL4.3 wild type Vif. It is worth noting that in our previous work (Bergeron, Huthoff *et al*. 2010), the ΔPPL triple-mutation construct had a significantly deceased ability to recruit the ubiquitination complex. Taking this comparison and previous biophysical studies (Simon, Sheehy *et al*. 1999) into consideration, we argue that 161PPL163 is essential for Vif function whereas a single amino acid change does not undermine Vif functions significantly.

As for the P164S mutant, no Vif protein was detected in the elution, indicating that the third proline plays an important role in forming the ubiquitination complex. In Dr. Julien Bergeron's work, the construct APLA, a double-point mutant, lost its ability to bind to the complex, and had a lower expression yield as well. Since the first proline does not affect the binding as we observed in figure 6.4, it is likely that the third proline is essential for forming the ubiquitination complex. Future work should address and clarify the biological role of the third proline.

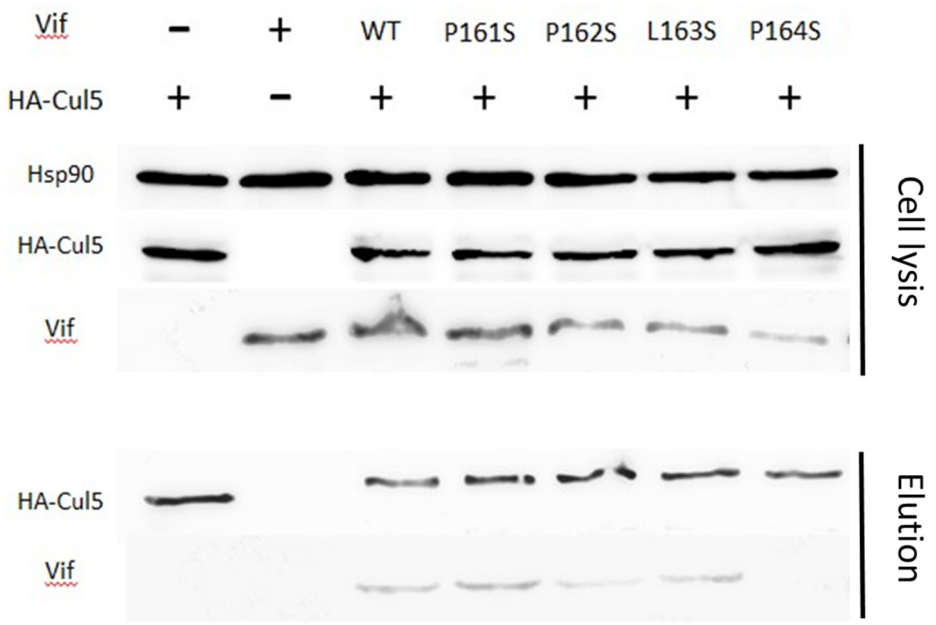


Figure 6.4: Co-immunoprecipitation study on the recruitment of ubiquitination complex. 293T cells were transfected with plasmids expressing WT Vif or Vif mutants and HA-tagged Cul5. HA-Cul5 was isolated by elution from anti-HA beads with HA peptides. Cell lysis and elution were analysed by western blot using anti-Vif or anti-HA antibodies.

6.3 Summary

In previous work the PPLPS motif of Vif had been studied, but it was largely still unclear how each residue contributes to the functions of the motif (Donahue, Vetter *et al*. 2008, Bergeron, Huthoff *et al*. 2010, Walker, Khan *et al*. 2010). In previous chapters I presented the structure of SOCS-box in the complex with EloBC and found that a single residual mutation did not affect the binding to EloBC by biophysical techniques. Here, the importance of residues L163 and P164 was confirmed whereas P161 and P162 do not seem to have crucial functions, at least not in the assays performed in the thesis.

Further experiments are required to understand the interface between the PPLPS motif and Cul5. Actually, the function of L163 is different from the function of P164 according to our biological assays. Mutation of the leucine decreased viral infectivity but did not disable Vif's ability to recruit the ubiquitination complex, which also implies that this leucine (L163) may be related to the interaction with A3G, whereas the mutation on the third proline removed Vif from the complex. Therefore, it is worth knowing whether the PPLPS motif of Vif, and the DVMK stretch of EloB will form a common interface for the recruitment of Cul5 by the interaction between Vif and EloB before further binding to A3G. It is possible that the PPLPS motif of Vif is the hub within the final Vif-A3G-Cul5-EloB complex.

Our results also imply that circulating HIV-1 strains may express Vif proteins with different anti-A3G activities as judged by performing single point mutations in the Vif from NL4.3. However, these results have to be interpreted very carefully and there are a couple of reasons why point mutants in NL4.3 Vif would lose anti-A3G activity or become instable, while these Vif changes may exist in circulating HIV-1 strains. It is possible that unidentified secondary compensating mutations are present in these strains and that Vif has kept stability and anti-A3G activity. Alternatively, polymorphisms may be present in Cul5 which may exert a selective pressure on HIV-1 Vif to change critical residues such as L163 or P164. Further studies should combine these ideas and look at the possibly flexible interplay between the host protein Cul5 and HIV-1 Vif.

CHAPTER 7: TOWARDS A PROTOCOL FOR SOLUBLE VIF

Chapter 1 discussed that Vif and Vif-related interactions are potential targets for the development of new anti-HIV drugs. Attempting to purify and crystallize Vif is the final objective of this Vif project. Until now, due to the insolubility of Vif, only the structure of Vif BC-box was achieved. The discovery of a chaperone in A3G degradation pathway, named CBFβ (isoform I), highlighted the possibility of obtaining soluble Vif from a prokaryotic system, e.g. from *E. coli*. In this Chapter efforts were made to search for a condition of expressing and purifying Vif from *E. coli*, and the entire process of obtaining a soluble Vif is described.

7.1 Vif sequential analysis

Initially, a bioinformatic analysis to characterize the Vif amino acid sequence was performed. The secondary structure predication by PsiPred shows that Vif is composed of an N-terminal β-strand, four short β-strands in the middle of sequence, four α-helices and a C-terminal coil tail (figure 7.1A). Interestingly, hydrophobic analysis indicates that most of the amino acid residues in the Vif sequence are hydrophilic (figure 7.1B), giving the value of the 'grand average of hydropathicity index' (Kyte and Doolittle 1982) as -0.734 which indicates a hydrophilic character. The value of the 'instability index II' (Guruprasad, Reddy *et al*. 1990) is 37.35 which means the protein is stable, whereas in experimental practice Vif is insoluble and easily aggregates. These conflicting results from theoretical predictions and practical experiments suggest that those hydrophobic residues may be exposed on the protein surface in structured Vif, or that the folding process of Vif during expression needs extra factors to help its folding. Therefore, we managed to co-express Vif with the chaperone CBFβ in *E. coli* since the chaperone was found to be critical in the folding and stabilization of Vif and Vif-mediated evasion of host immune responses (Jager, Kim *et al*. 2012, Zhang, Du *et al*. 2012).

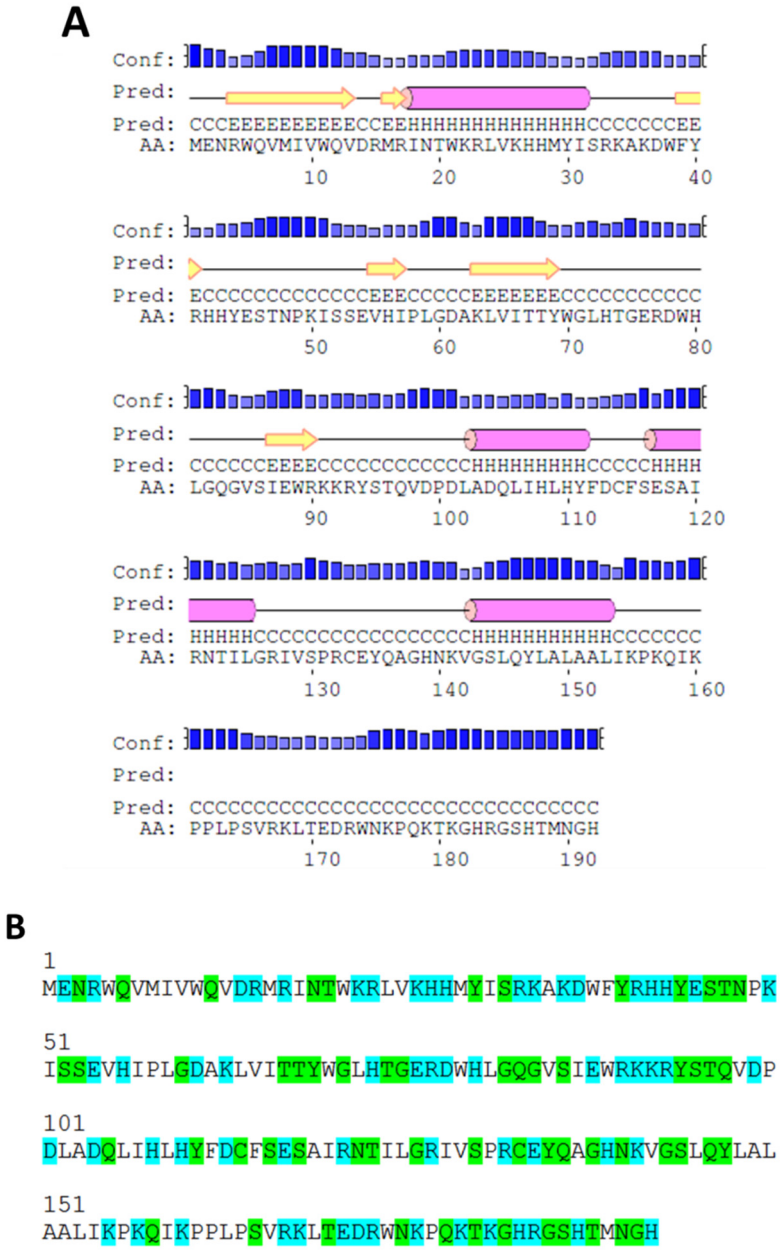


Figure 7.1: Vif protein sequence analysis. (A) The secondary structure prediction on Vif was obtained using PsiPred. (B) Hydrophobicity analysis was performed on Vif sequence.

7.2 Expression test of Vif

7.2.1 Co-expression of Vif with core binding factor β

In the first step, we used a truncated CBFβ (tCBFβ) construct (residues 1-140) rather than a full-length construct for three reasons. First, there are two isoforms of CBFβ, sharing over three-quarters identity of the N-terminal amino acid sequences. On the other hand, the molecular weight of full-length CBFβ is very close to that of Vif (22 and 24 kDa. respectively), which means they may not be separated completely on a gel, which may influence the quantitative analysis and judgment just by Coomassie blue. Besides, the structure of the core body of CBFβ (residues 1-140) has been solved (Goger, Gupta *et al*. 1999), confirming the importance of the N-terminus. Therefore, a tCBFβ construct was used as a starting construct as this facilitated the preliminary studies from which we might know whether tCBFβ was able to help Vif fold properly in bacteria or a full-length construct was required.

Vif and tCBFβ were co-expressed in LB media at 37 ℃ or 18 ℃, followed by re-suspending, sonication and spinning cell debris down in Tris buffer with different pH and a gradient concentration of NaCl. The SDS-PAGE gel shows that if proteins were expressed at 37 ℃, Vif protein remained in the pellet, termed inclusion bodies, after a hard spin regardless of buffer conditions. Besides, at 37 ℃ tCBFβ was expressed in larger amounts than Vif (figure 7.2A, B), which was probably caused by the high solubility of CBFβ. Noticeably, if proteins were expressed at low temperature, the Vif band was still visible in the supernatant lane on the SDS-PAGE gel after a hard spin (figure 7.2C), showing that Vif only folded properly at low temperature, and that salt concentration and pH impacted on the stability of Vif in solution (figure 7.2D). Meanwhile, this result acquired from tCBFβ construct suggests that the N-terminal 140 residues are sufficient for Vif folding in the presence of CBFβ, which was also confirmed by a recent studies (Kim, Kwon *et al*. 2013).

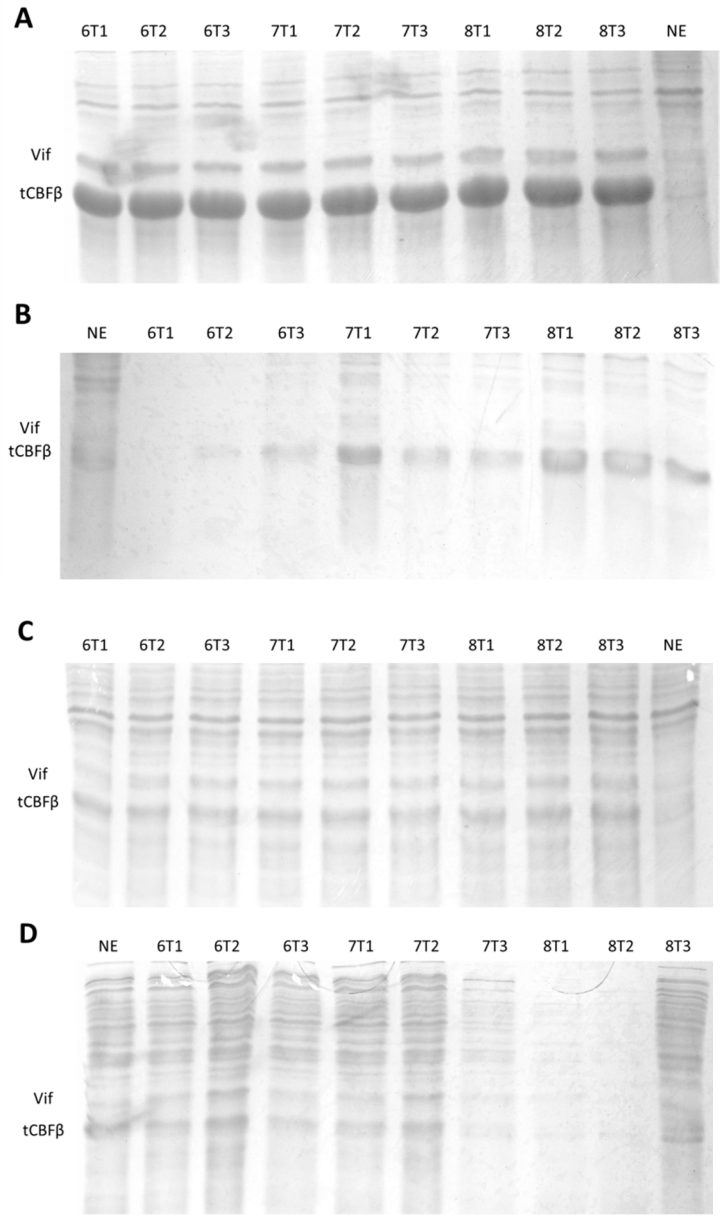


Figure 7.2: Expression test of Vif co-expressed with truncated CBFβ. (A) Cell lysis of samples at 37 ℃. (B) Supernatant of cell lysis from 37 ℃. (C) Cell lysis from 18 ℃. (D) Supernatant of samples from 18 ℃. NE and T stands for non-expression control and Tris buffer. 6, 7 and 8 stands for the pH. 1, 2, 3 indicates NaCl concentration 50, 250 and 500 mM respectively.

7.2.2 Co-expression of Vif, Elongin B, Elongin C and core binding factor β

In the next step we managed to co-express Vif His_6-fused EloB, EloC, with CBFβ isoform I, the isoform that other groups have used in their published work (Jager, Kim *et al*. 2012, Zhang, Du *et al*. 2012). In order to achieve that Vif and CBFβ sequences were cloned into a pRSFDuet vector with a kanamycin antibiotic manner, and EloB and EloC sequences onto a pETDuet vector (constructed by Dr. Julien Bergeron) with an ampicillin antibiotic manner. The two plasmids were co-transformed into *E. coli* BL21 (DE3) pLysS competent cells selected for resistance to the antibiotics (figure 7.3A).

Similarly, the proteins were expressed at 18 ℃ or 37 ℃ but also with a gradient of IPTG concentration. Cells were dealt with in the same way as in the previous experiments in Tris buffer, pH 7.0, 250 mM NaCl. Supernatants were loaded onto Tricine gels for analysis. Figure 7.3B reveals that at lower temperature with lower IPTG induction concentration Vif was more soluble and stable in solution. This is because low temperature and less IPTG decreased the rate of protein production, providing more time for the protein to fold. Besides, the yield of individual protein was very closer to a 1:1:1 stoichiometry whereas at higher temperature more EloBC dimer was expressed. Therefore, it is now known that soluble Vif in the presence of CBFβ, EloB and EloC can be obtained by co-expression in LB media supplemented with low IPTG concentration at low temperature.

7.3 Solubility test for Vif

7.3.1 Solubility test for Vif in the tetramer

In order to further confirm that Vif remains in the supernatant and to find the best buffer condition for extracting Vif from cells, we co-expressed the four proteins and re-suspended them in different buffers as in previous experiments. Because of the similar molecular weight of CBFβ and Vif, these two proteins could not be separated completely on a SDS-PAGE gel, and the difference in band intensity among different groups is not clearly distinguished on the

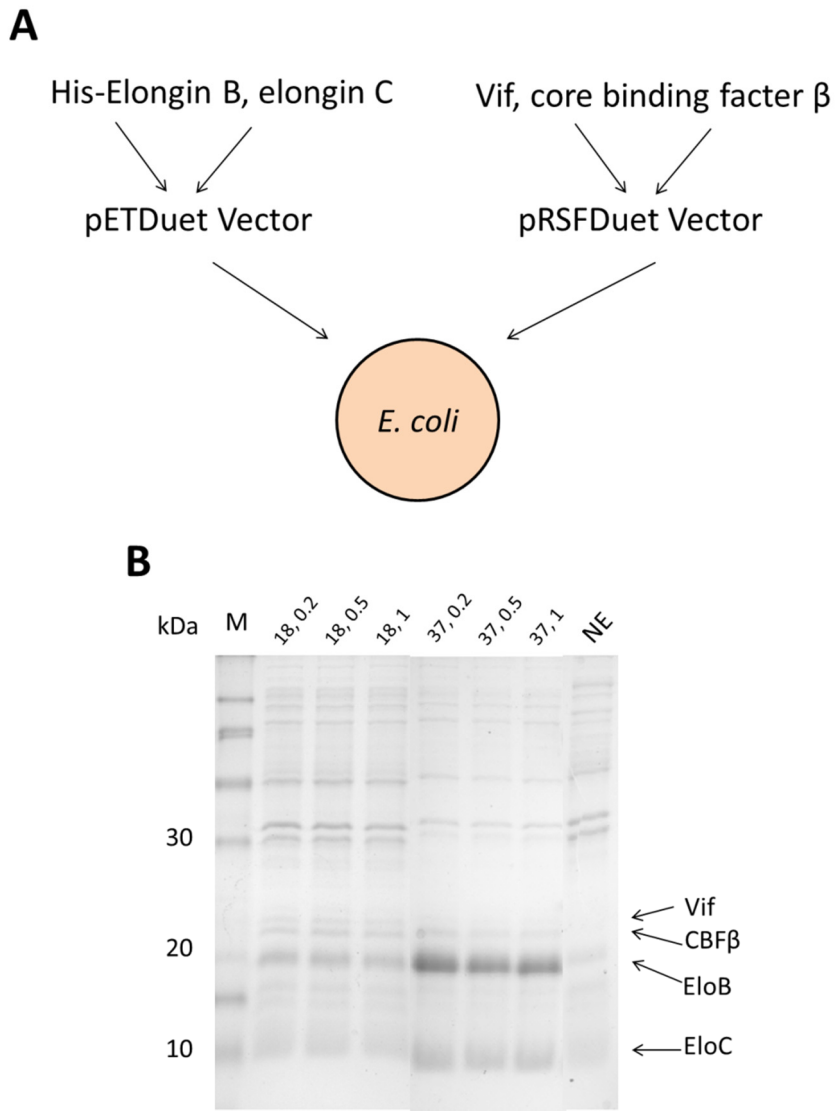


Figure 7.3: Expression test of Vif co-expressed with CBFβ/EloBC. (A) Schematic representation of co-expression strategy of the tetramer. EloB and EloC CDs were cloned onto pETDuet Vector. Vif and CBFβ CDs were constructed onto pRSFDuet Vector. Two vectors were co-transformed into *E. coli*. (B) Expression test of the tetramer at body temperature or at cold temperature with various IPTG concentration. 0.2, 0.5 and 1 indicates the mM of final IPTG concentration. NE stands for non-expression control.

gel (figure 7.4A). Thus we used western blotting to detect Vif protein in each sample using anti-Vif primary antibody. As was expected, the immunoblotting demonstrated that the target protein was present in the supernatant. It also reveals that Vif is more soluble at a low salt concentration at pH 7.0-8.0 in the sonication and spinning steps (figure 7.4B).

Then we managed to compare various types of buffer by the same method. The immunoblot provided the same trend for the decreasing amount of Vif in the supernatant along with the rising salt concentration at pH 7.0 (figure 7.4B). Interestingly, Vif in the supernatant with sodium-phosphate buffer was not affected by salt concentration (figure 7.4C), suggesting that phosphate buffer is the best choice to keep Vif stable when Vif was separated from cells.

7.3.2 A study on Elongin BC and core binding factor β interaction

Now it is possible to purify Vif together with His_6-EloBC and CBFβ from supernatant on a Ni-NTA column according to the previous results. However, considering that there was only one His_6 tag on EloB component, we first asked whether or not CBFβ interacts with EloBC. As it is known that Vif binds to EloBC and CBFβ respectively, if CBFβ does not bind to EloBC, and meanwhile if four bands from the elution fractions are observed on the gel, it will be undoubtedly that soluble Vif is purified successfully by this co-expression and purification strategy.

Following this assumption we prepared purified CBFβ and EloBC individually and mixed them together (figure 7.5A). This mixture was then loaded on to a sizing column (Superdex 75 10/300 GL). Fractions were collected and loaded onto gels for analysis. Figure 7.5B and C appears that there is no binding between EloBC and CBFβ. These two proteins were separated completely according to the gel filtration analysis. This result indicates that untagged CBFβ cannot be acquired by Ni-NTA purification in the presence of His-EloBC whereas in the absence of Vif.

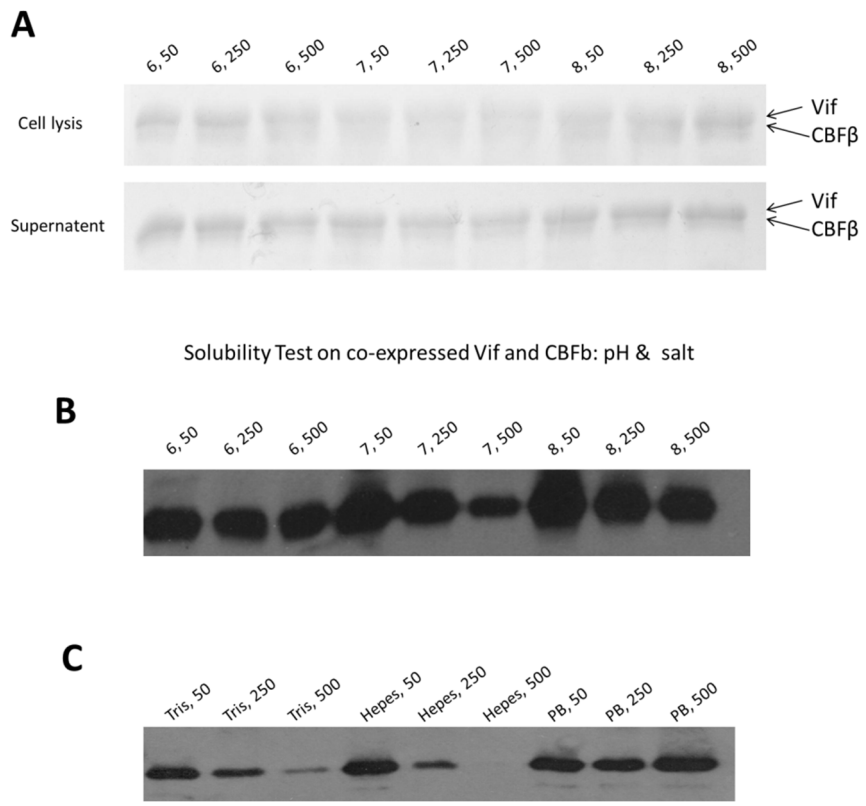


Figure 7.4: Solubility test for Vif in the tetramer. (A) SDS-PAGE analysis of pH & salt test in Tris buffer. 6, 7, 8 stands for the pH and 50, 250, 500 indicates 50, 250, 500 mM of NaCl. (B) Western blot of pH and salt tests. (C) Western blot of buffer tests. Anti-Vif antibody is used to detect Vif in the samples. PB stands for phosphate buffer.

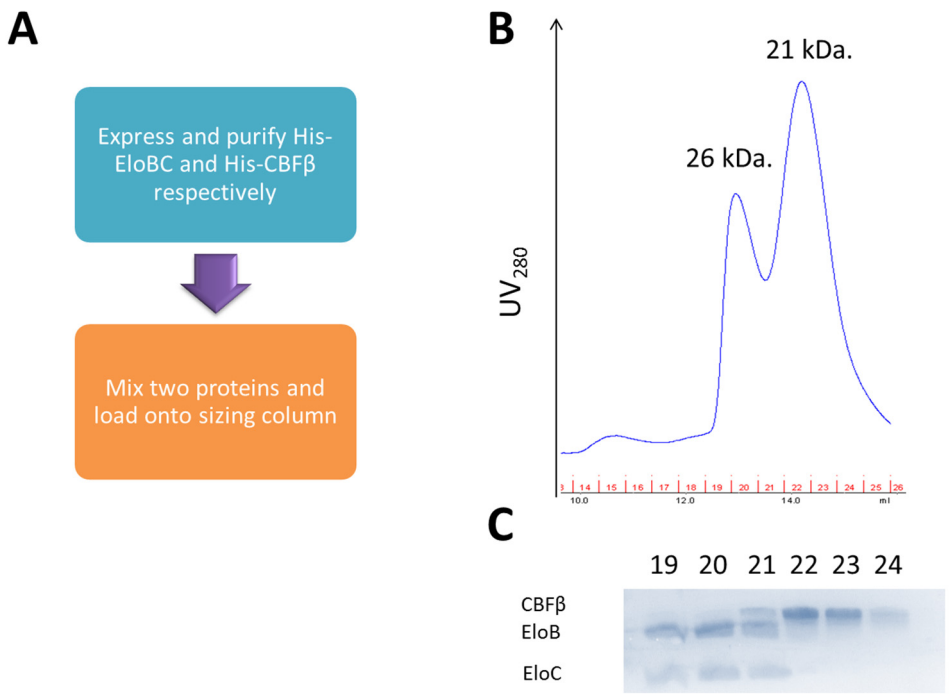


Figure 7.5: EloBC and CBFβ interaction study. (A) Schematic representation of this assay. (B) The EloBC-CBFβ mixture was loaded onto a sizing column and detected by UV280. Two peaks were observed. (C) SDS-PAGE gel analysis of fractions 19-24.

7.4 Stability test for Vif after purification

7.4.1 Detergent test

In order to continue the search for the best buffer conditions for Vif in the Vif/CBFβ/EloB/EloC tetramer, several detergents were tested. Briefly, after sonication, the cell lysis was distributed into several tubes. Appropriate detergents were added to each tube, followed by high-speed centrifugation and purification from Ni-NTA columns. Water was used as a negative control, i.e. no detergent (figure 7.6A). The Tricine gel shows that Tween 20, NP40 and CHAPS improves the final yield of the target proteins compared to the negative control whereas Triton, a widely used detergent in protein preparation, is harmful to Vif in solution (figure 7.6B). The protein concentration in the final fractions was measured immediately after elution or after three days. It suggests that 90% of the proteins stayed in solution with CHAPS, and that all the proteins came out of the solution with Triton after three days. The rest of conditions kept between 80% to 88% of proteins in their individual condition (figure 7.6C), indicating that CHAPS can improve the amount of soluble Vif and keep Vif stable under artificial buffer conditions.

7.4.2 Hunting for conditions to enhance stability

Finally the Thermal Shift Assay (TSA) was applied to optimize conditions accurately for the stability of purified Vif. Basically, in each well of a 96-well PCR plate there is a mixture of fluorescent dye, testing buffer and approximately 0.02 mg/ml of protein. At the beginning the dye is quenched in a hydrophilic environment, i.e. buffer. The protein then starts denaturing with increasing temperature by qPCR machine, thus exposing its hydrophobic residues which will take and activate the dye. The machine records the signals at each temperature level quantitatively. Under favourable conditions protein denatures at a higher temperature whereas under poor conditions signals can be detected at a lower temperature. The temperature at which around 50% of the proteins denature is taken as an index, namely T_m, the higher the better.

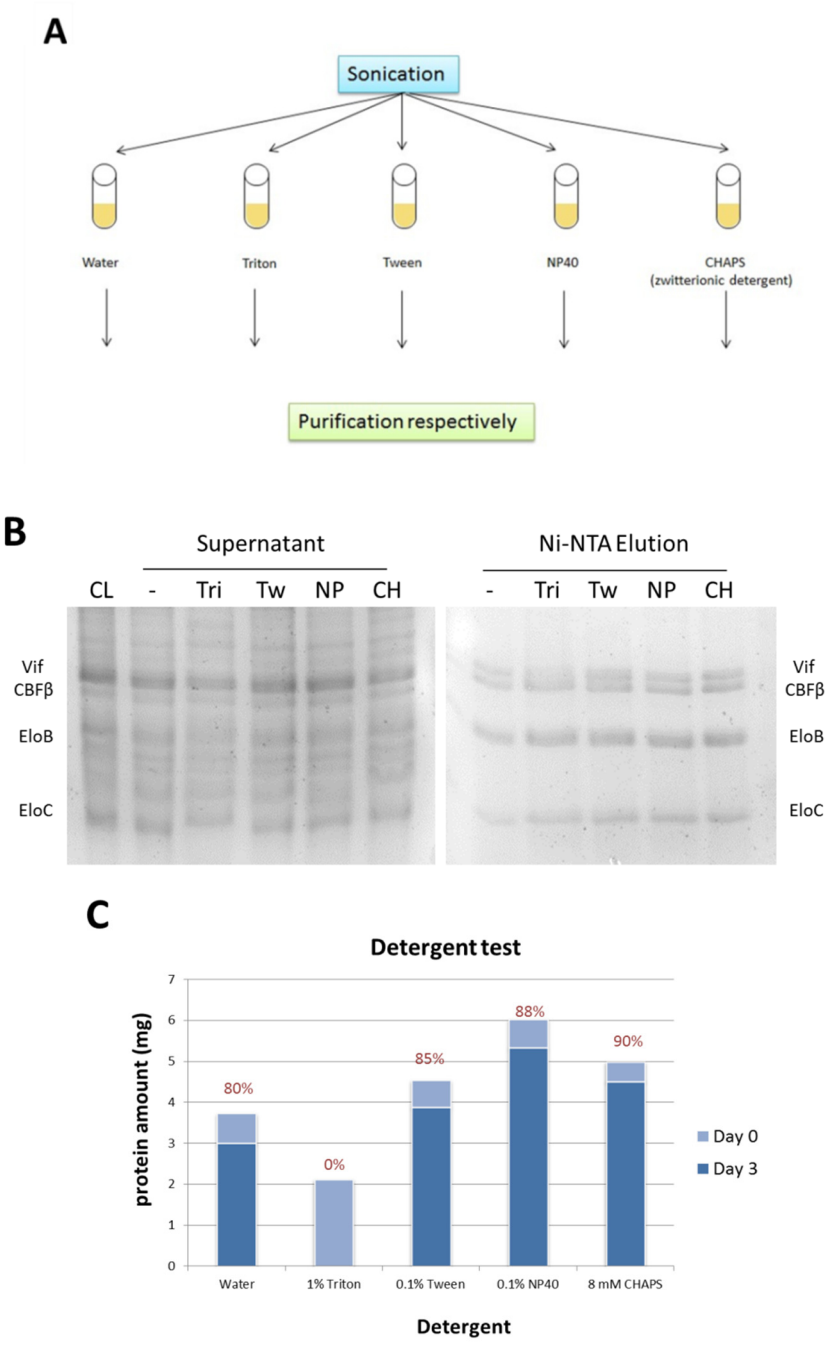


Figure 7.6: Detergent hunting for Vif stability. (A) Schematic representation of this test strategy. In this case 20 mM Tris buffer, pH 7.0 with 250 mM NaCl was used as a basic buffer. (B) Tricine gel analysis of purification with different detergents. Left, supernatant. Right, elution from Ni-NTA column. (C) The yield of protein was measured by BioRad kits immediately after elution and on the third day. Percentage stands for the amount of protein left in supernatant after a hard spin. CL, -, Tri, Tw, NP, CH stands for cell lysis, no detergent, Triton-20, Tween, NP40 and CHAPS respectively.

Figure 7.7A shows that in HEPES buffer protein samples prefer acidic conditions with an ionic strength in the mild range. Similarly figure 7.7B reveals that in phosphate buffer protein samples prefer alkaline conditions with a mild ionic strength. It is of note that in this assay the mild ionic strength is the best option for Vif whereas in the solubility test low salt concentration is the best (figure 7.4B). This is because the solubility test was performed within cell lysis and supernatant, which means endogenous ions from cells actually improved the ionic strength in the conditions regardless of external 50 mM salt. On the other hand, this TSA assay was performed on purified proteins and the ionic strength was only provided by the salts in the buffer. Thus these two results provide clear information that Vif prefers mild salt concentration. Meanwhile it is observed that the T_m in all phosphate buffer conditions are comprehensively higher than that in HEPES conditions, which agrees with the previous result from immunoblotting described above (figure 7.4C), confirming that phosphate buffer is the best one for Vif. In addition, we compared the effect of NaCl and KCl by the same technique. Figure 7.7C reveals that NaCl is better than KCl at pH 7.0. Therefore, it now can be concluded that the best buffer for Vif when in the tetramer is 20 mM phosphate buffer, pH 7.0 containing around 250 mM NaCl and 8 mM CHAPS.

7.5 Understanding the characteristics of CBFβ

7.5.1 Vif/CBFβ and EloBC pull-down assay

As soluble Vif can be obtained by the methods detailed above, it is interesting to know whether CBFβ helps Vif fold properly in order to obtain biological functions with or without EloBC. A pull-down assay was then performed to address this question. Briefly, untagged Vif and untagged CBFβ were co-expressed. The cell lysis after sonication was spun down and the supernatant was loaded onto a Ni-NTA column on which His_6-tagged EloBC was bound (figure 7.8A). Proteins were eluted by approximately 400 mM imidazole after intensive wash. Cell lysis, supernatant, flowthrough and elution fractions were taken and mixed with loading buffer for Tricine gel analysis.

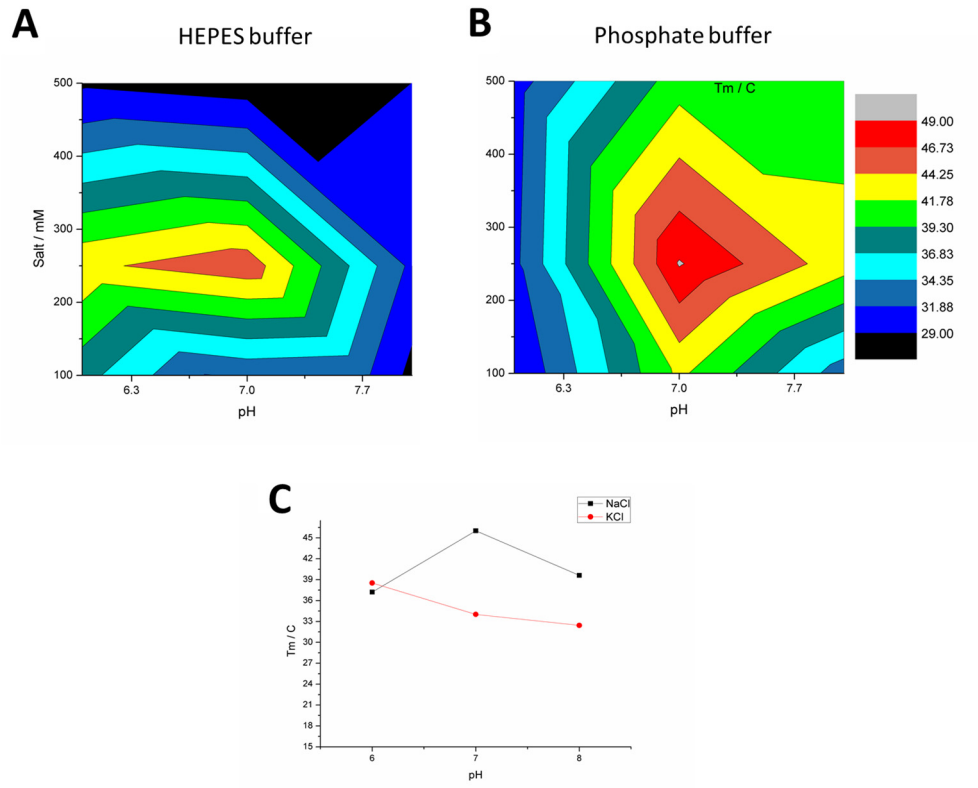


Figure 7.7: Hunting for conditions for Vif stability by thermal shift assay. (A) Salt and pH optimization in HEPES buffer. (B) Salt and pH optimization in phosphate buffer. (C) The effect of NaCl and KCl on protein stabilization.

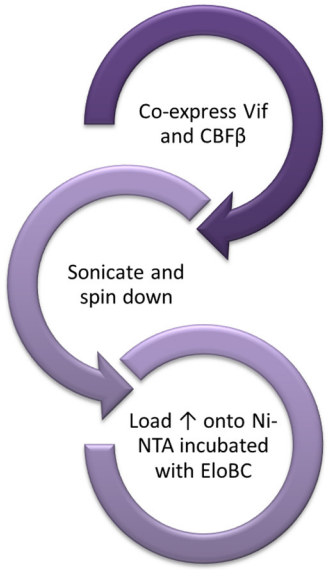
Figure 7.8B reveals that Vif and CBFβ bands disappear in the flowthrough lane compared with the cell lysis and supernatant lanes, whereas in the elution lane two bands are clearly visible besides the EloBC bands. As it is known that CBFβ does not bind to EloBC, these two new bands are definitely from Vif and CBFβ respectively, confirming that in our study CBFβ is the only crucial factor that is required to help Vif fold in *E. coli* so that it attains its functional structure.

7.5.2 A study on CBFβ isoform II

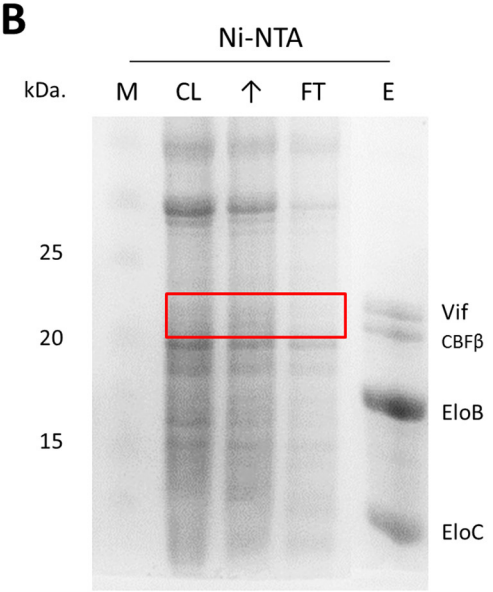
In the previous studies we focused on CBFβ isoform I. Now we want to understand whether CBFβ isoform II has the same function as isoform I does. Sequential alignment shows that these two isoforms share a 90% identity of amino acid sequences. Only the carboxyl terminal residues are different (figure 7.8C). The published solution structure (PDB ID: 1CL3) reveals that the core body of CBFβ (residues 1-140) is a stable barrel-like structure composed of 6 β-sheets (Goger, Gupta *et al*. 1999) whereas the structure of full-length CBFβ has not been reported. Therefore, if the behaviour of isoform I and isoform II is found the same, it will indicate that this core body is the only region that is crucial for Vif folding as well as genetic regulation, its original function.

Similarly, Vif, CBFβ isoform II and EloBC were co-expressed with low ITPG concentration (0.2~0.25 mM) at low temperature (16~18 ℃), followed by purification on a Ni-NTA column and a sizing column. Interestingly, the Tricine gel displays that the yield of EloBC was much higher than the yield of Vif and CBFβ isoform II (figure 7.8D). It also shows that in the first step some of Vif was lost in the inclusion bodies due to the unfolded aggregation and in the flowthrough due to incorrect folding that disables the binding to EloBC. The functional difference between two CBFβ isoforms intimates that the carboxyl terminus on isoform I participates in facilitating Vif to fold during the expression.

A



B



C

Isoform 1

MPRVVPDQRS KFENEEFFRK LSRECEIKYT GFRDRPHEER QARFQNACRD GRSEIAFVAT
GTNLSLQFFP ASWQGEQRQT PSREYVDLER EAGKVYLKAP MILNGVCVIW KGWIDLQRLD
GMGCLEFDEE RAQQEDALAQ QAFEEARRRT REFEDRDRSH REEMEVRVSQ LLAVTGKKTT
RP

Isoform 2

MPRVVPDQRS KFENEEFFRK LSRECEIKYT GFRDRPHEER QARFQNACRD GRSEIAFVAT
GTNLSLQFFP ASWQGEQRQT PSREYVDLER EAGKVYLKAP MILNGVCVIW KGWIDLQRLD
GMGCLEFDEE RAQQEDALAQ QAFEEARRRT REFEDRDRSH REEMEARRQQ DPSPGSNLGG
GDDLKLR

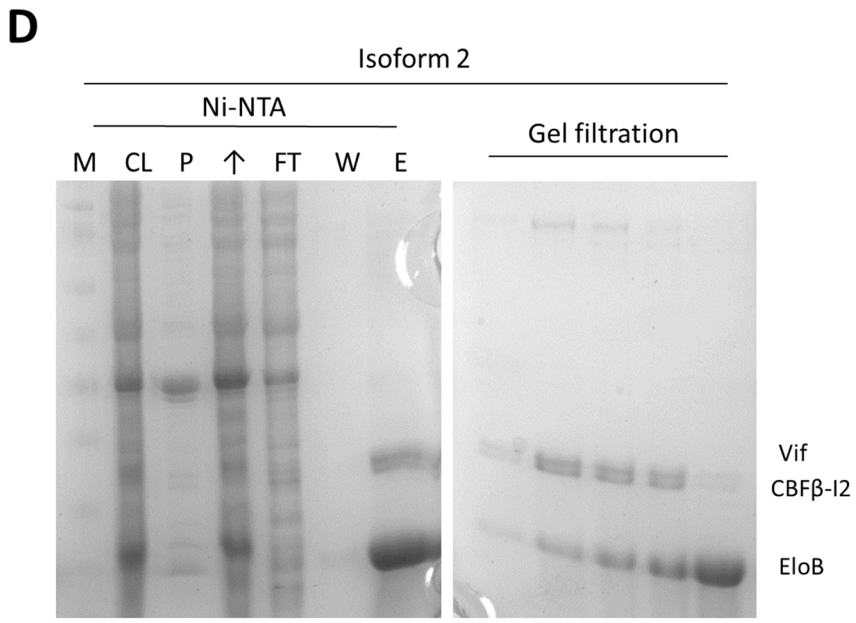


Figure 7.8: The characteristics of CBFβ during Vif expression process. (A) Schematic representation of the pull-down assay. (B) Tricine gel analysis of the pull-down assay. The red square indicates that Vif and CBFβ were not left in flowthrough. M, CL, ↑, FT and E stands for marker, cell lysis, supernatant, flowthrough and elution respectively. (C) The sequences of two CBFβ isoforms. The carboxyl tail is presented in different colours. (D) Vif was co-expressed with EloBC and CBFβ isoform 2. Tricine gel analysis was used to check Vif behaviour. M, CL, P, ↑, FT, W and E stands for marker, cell lysis, pellet, supernatant, flowthrough, washing and elution respectively.

7.6 Summary

In this chapter the attempts which were made to acquire purified soluble Vif protein for biochemical and structural studies are described. In a prokaryotic expression system, Vif protein is able to be extracted from cells in the presence of a chaperone CBFβ isoform I. Owing to the solubility and stability of Vif, it must be expressed at low temperature with low levels of IPTG induction. The purification process must be conducted at low temperature as well. Zwitterionic detergent CHAPS is essential for Vif stability under the artificial conditions. The intensive studies made it possible to set up the crystallization screen, which will be discussed in the next Chapter.

On the other hand, CBFβ isoform II was investigated. Differing from isoform I, isoform II contains a weak capability to help Vif fold in cells, because of which more EloBC was expressed rather than the tetramer of interest. This result implies that the C-terminus of isoform I is important for Vif folding. The last Chapter will discuss probable future work to identify the structural mechanism.

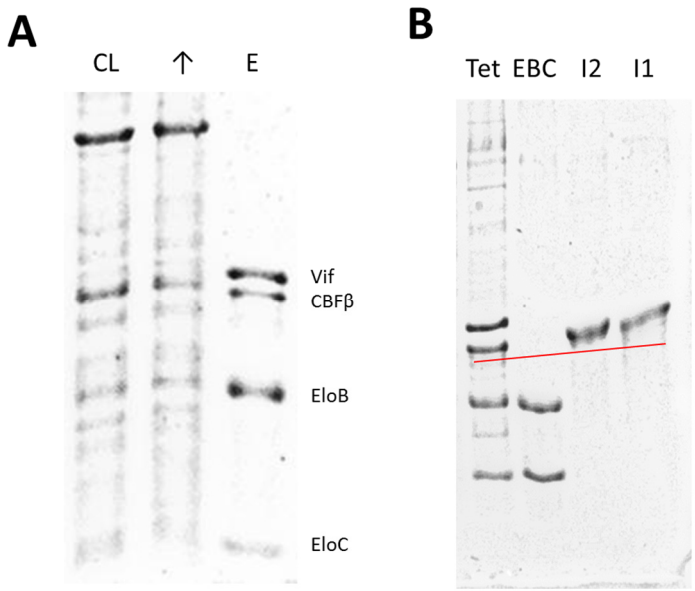
CHAPTER 8: CRYSTALLIZATION OF VIF

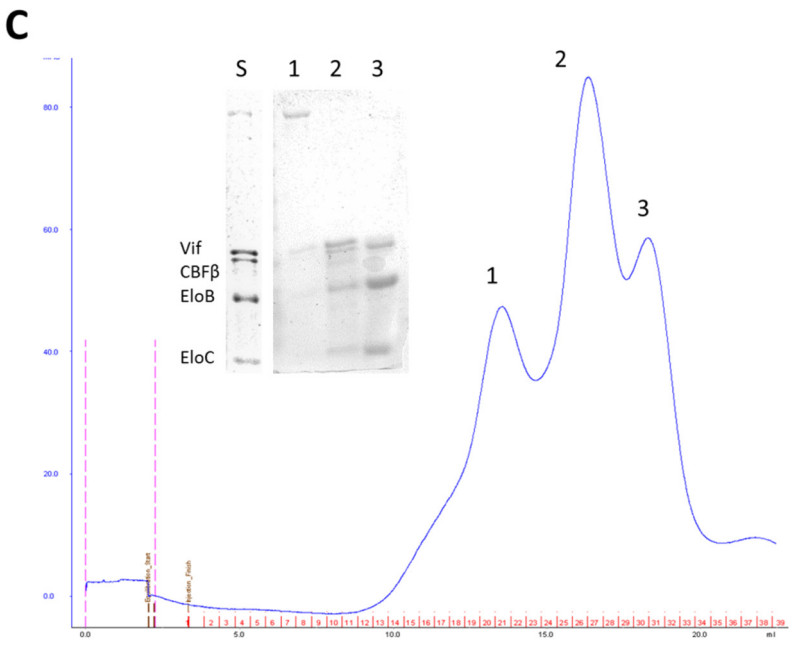
In the previous Chapter 1 discussed how soluble Vif in the Vif/CBFβ/EloB/EloC hetero-tetramer was acquired after employing many biochemical and biophysical assays in order to find a suitable expression and purification strategy. This Chapter continues to describe the complete protocol for acquiring purified Vif. Subsequently efforts were made to crystallize the hetero-tetramer (tetramer) for structural studies by X-ray crystallography.

8.1 The preparation of Vif/CBFβ/EloBC complex

The four components of the complex were co-expressed and co-purified in the manner detailed in Chapter 2 and Chapter 7. The Tricine gel (figure 8.1A) shows how an affinity chromatographic purification followed by intensive washing achieves a sample of a very high purity level (figure 8.1A). It is worth mentioning that sometimes there is another band or a small region of smear just below the second band (the CBFβ band). In order to identify the bands, the first band, the second band and the unclear third band were cut out respectively for mass spectrometry analysis. The results from MS confirmed the presence of Vif in the first band and the constituent present of molecules in the second band and the third small band (raw data are included in the electronic version). In addition, CBFβ monomer and EloBC dimer were prepared and checked by Tricine gel respectively. The gel reveals that the smear or the other small band below the second band might be only caused by a strange or aberrant mobility of CBFβ in gels. The gel also demonstrates that the first clear band at the top is clearly a Vif band (figure 8.1B).

In the next step the elute from the nickel column was loaded onto a gel filtration column washed with the final buffer (10 mM Na-phosphate buffer, pH 7.0, 0.02% NaN_3, 8 mM CHAPS, 250 mM NaCl) to remove the excessive EloBC dimer. Figure 8.1C is a typical profile from the gel filtration column. Usually there are three clearly visible peaks in each run. The first peak coming out very early in the figure was composed of an aggregate of Vif and large-molecular


A
CL
↑
E
Vif
CBFβ
EloB
EloC
B
Tet EBC I2 I1


C
S 1 2 3
Vif
CBFβ
EloB
EloC
1
2
3

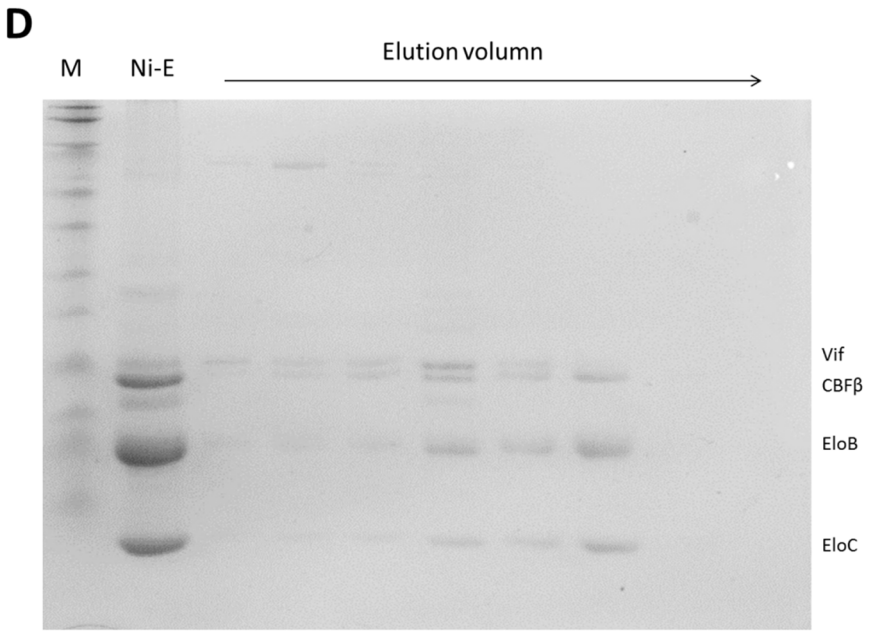


Figure 8.1: The profile of the tetramer purification. (A) Tricine gel analysis of the purification on Ni-NTA column. (B) Complex, EloBC and two CBFβ isoforms were loaded in one gel to validate Vif band. CL, ↑, E, Tet, EBC, I2, I1 stands for cell lysis, supernatant, elution, tetramer, EloBC, isoform 2, and isoform 1 respectively. (C) The protein samples eluted from Ni-NTA were loaded onto a gel filtration column for further purification. Tricine gel analysis was performed to see proteins from each peak. S, 1, 2, 3 stands for samples before loading, peak 1, 2 and 3 respectively. (D) Complete Tricine gel analysis of the sizing column. Ni-E means elution from Ni-NTA.

impurities. The target protein complex was eluted in the second peak after approximately 14 ml of washing volume. Extra EloBC (26 kDa.) and disassociated CBFβ (22 kDa.) from Vif aggregate were separated from the tetramer and came out in peak 3. Figure 8.1D shows that the changing of composition within each fraction during the sizing column purification. The positions of each peak and the corresponding molecular weights match those from a group of protein ladders used for column calibration (table 2.2). It confirms that the protein in peak 2 is the tetramer of interest with approximately 71 kDa molecular weight.

8.2 High-throughput screening from Hauptman Woodward Institute

Hauptman Woodward Medical Research Institute (HWI) at Buffalo, New York is a famous institute renowned for its structural biological studies (http://www.hwi.buffalo.edu). The high-throughput screening laboratory (HTS lab) at HWI is equipped with a robust crystallization trial that tests 1536 conditions in one set of experiments with standard screen kits (Luft, Collins *et al*. 2003) or membrane screen kits (Koszelak-Rosenblum, Krol *et al*. 2009). These kits were constructed based on the results from a large number of trials performed at the HTS lab. These data are integrated into a crystallization database that is used to produce crystallization recipes. Photographs for each well, i.e. condition, are taken on day 0, day 1, week 1, 2, 4 and 6, which can be used for the analysis when crystals grow.

The final tetramer sample sitting in the final buffer was concentrated to 4 mg/ml in 700 µl volume as measured by BioRad protein assay kit and was sealed in a falcon tube placed in a thermos. A styrofoam box containing the sample in a thermos with ice bags was shipped to the HTS lab in four days. Samples were set up immediately at 14 ℃ when they arrived.

Unfortunately, the protein precipitated in the majority of the conditions in less than one week. However, in several conditions there was something special that could be observed after two weeks as shown in figure 8.2. Basically, tree-like crystals grew in conditions containing over 0.1 M $MgCl_2$. These kinds of crystals are possibly salt crystals according to their morphology. In another two conditions (0.2 M $CaCl_2$, 0.1 M HEPES pH 7.0, 40% w/v

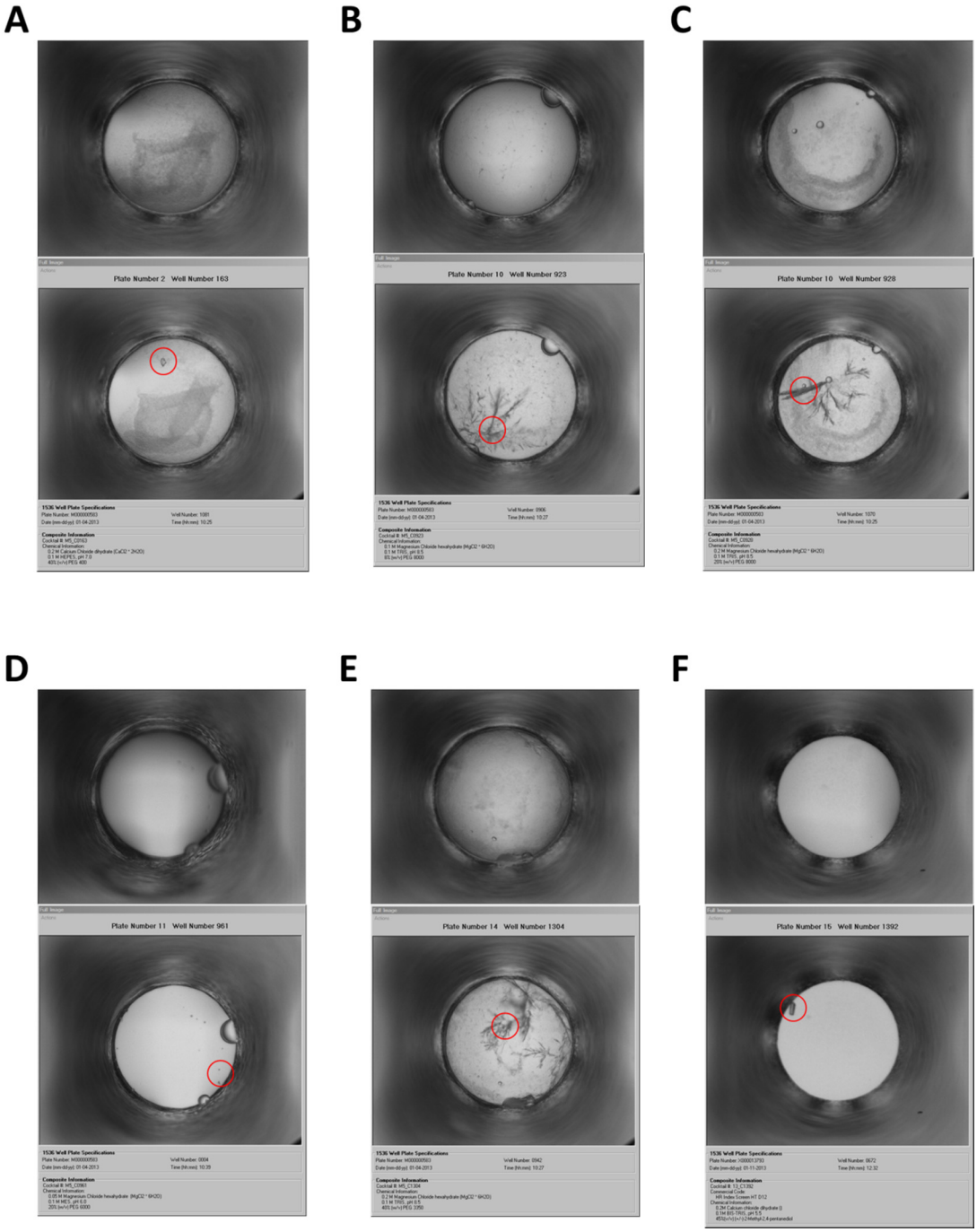


Figure 8.2: Crystallization results from the high-throughput screening lab. The changes of interest are circled. The up photographs were taken one day after setting. The lower photographs were taken four weeks after setting.

PEG400 and 0.2 M $CaCl_2$, 0.1 M Tris pH 5.5, 45% (+/-)-2-Methyl-2,4-pentanediol), some interesting particles were found, but they cannot be confirmed in the screening system because of technical limitations. Therefore, these conditions were repeated further and modified by hanging drop methods in-house, which will be discussed in the next section.

8.3 In-house screening for crystallization conditions

8.3.1 High-throughput screening using a Mosquito crystallization robot

The in-house crystallization trials were conducted using hanging or sitting drop methods. As it was described in Chapter 2, a Mosquito® robot was used to set 96-well plates for crystallization kit screening. In order to prevent the protein precipitating at high temperature in the middle of setting up, sample containers and plates with solutions were cooled down on ice for 15 min before placing on the machine. Plates were sealed and placed in a 4-℃ cold room immediately after setting up. Because the complex of interest consists of proteins, several kits that were developed for complex screening were tested in the past months. All kits which were tried are listed in table 8.1 below:

Table 8.1 The manufacturer's kits used for crystallization condition screening

Kit name	*Manufacturer*	*Description* [a]	*Comments*
Morpheus	Molecular dimensions	Derived from a database	Precipitation, S.C. [b]
Memsys		For transmembrane proteins	Precipitation
ProPlex		For protein complexes	Precipitation
JCSG-Plus		A highly effective strategy	Precipitation
Cryo I, II	Emerald Biosystems	Cryoprotectant conditions	Precipitation, S.C.

a. The main character of each kit is listed. The full details can be found on the manufacture website.

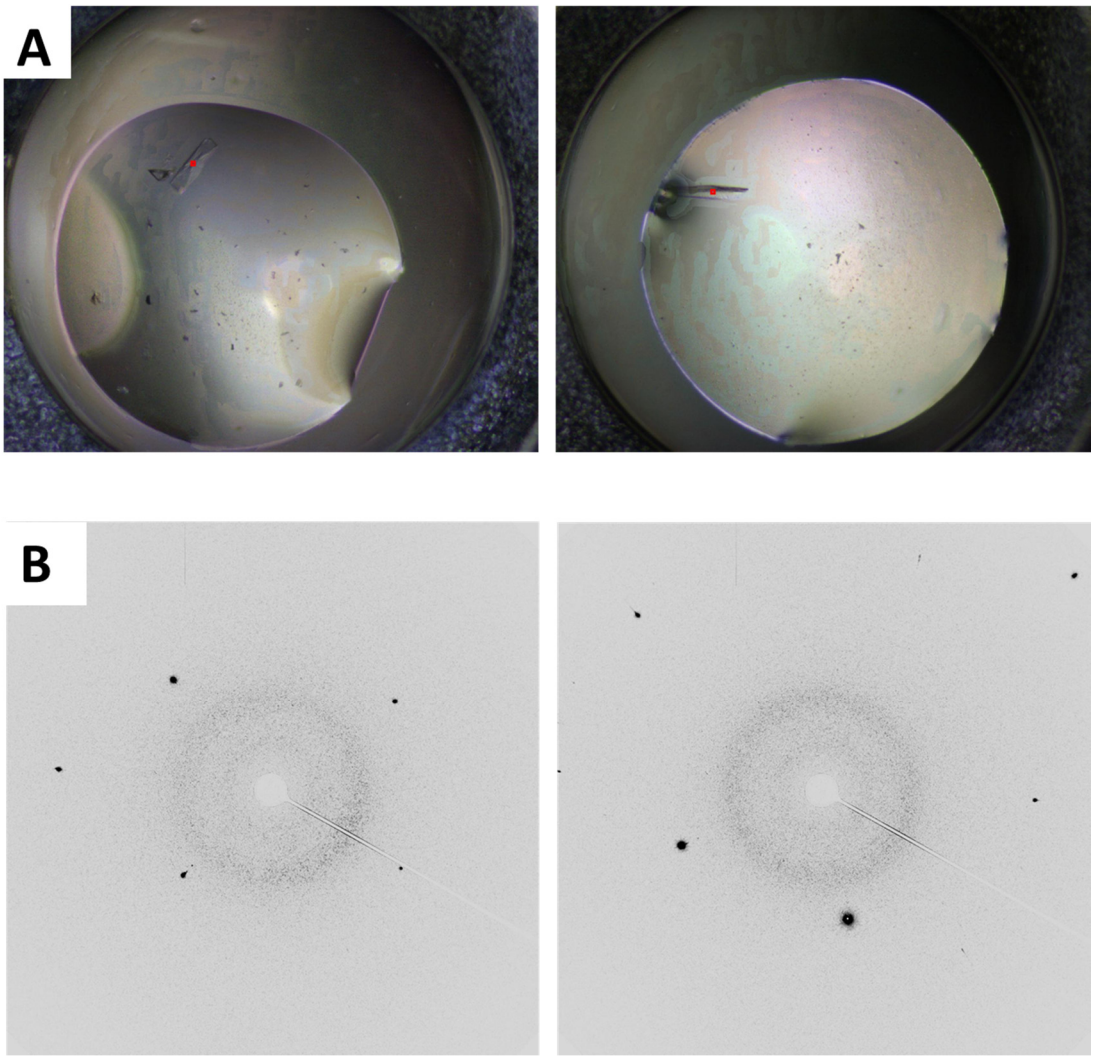
b. S.C. stands for salt crystals that were observed in some conditions from the kit.

The Morpheus kits were derived from an extensive data mining of over 33,000 PDB records, covering a range of pH, salt additives and PEGs. All the conditions in this set of kits are cryo-protected. After several days of setting up the trials, crystals were found in a few conditions. Figure 8.3A shows two conditions containing 0.09 M salt ($NaNO_3$, Na_2HPO_4, $(NH_4)_2SO_4$), pH 7.5 0.1 M Na-HEPES buffer with 4-Morpholinepropanesulfonic acid, 12.5% v/v 2-Methyl-2,4-pentanediol, 12.5% w/v PEG1000, 12.5% w/v PEG3350 and 0.09 M salt ($NaNO_3$, Na_2HPO_4, $(NH_4)_2SO_4$), pH 8.5 0.1 M Tris-Bicine, 20% v/v Ethylene glycerol, 10% w/v PEG8000 respectively. In order to test whether these were protein or salt crystals, the crystals were tested using the PX-scanner. The salt-detect method was chosen to shot the crystal. In theory, protein crystals give well-distributed, dense and small spots in the diffraction image, whereas salt crystals give only several strong spots in the image. Figure 8.3B shows the diffraction image acquired from the crystals displayed in figure 8.3A, suggesting that they were salt crystals rather than protein crystals. Because of the limited clear and strong spots in the image, the crystals were possibly from the sodium salts as Na^+ concentration in the solution was approximately 200 mM. Crystals from the other conditions were also checked and confirmed as salt crystals.

In the other kits from Molecular Dimensions only precipitation was observed. As for cryo kits from Emerald Biosystems, some crystals were obtained in conditions with a high concentration of salts and all of them finally were identified as salt crystals. These results imply that to crystallize Vif is still a tough challenge although it can be purified.

8.3.2 Screening by the hanging drop method

Samples of the complex were prepared and set up using the hanging drop method in-house as well as by the sitting drop method to mimic the conditions from HTS lab and to test for new conditions. The designed screening conditions were based on the phase diagram theory as elaborated in Chapter 2. Basically, a moderate concentration of precipitant was tested against pH gradients first. If the condition is still clear after several days, it means that higher concentration of the precipitants is required to push the complex into a nucleation zone (figure 2.5) and vice versa. By using this method a condition may be found where crystal will



8.3: Results from the screening in house. (A) Photographs on the salt crystals. (B) The diffraction images from the salt crystals. Left panel, 0.09 M salt ($NaNO_3$, Na_2HPO_4, $(NH_4)_2SO_4$), pH 8.5 0.1 M Tris-Bicine, 20% v/v Ethylene glycerol, 10% w/v PEG8000. Right panel, 0.09 M salt ($NaNO_3$, Na_2HPO_4, $(NH_4)_2SO_4$), pH 7.5 0.1 M Na-HEPES buffer with 4-Morpholinepropanesulfonic acid, 12.5% v/v 2-Methyl-2,4-pentanediol, 12.5% w/v PEG1000, 12.5% w/v PEG3350.

grow. The tested conditions and results of each condition are collated in table 8.2 (Presented in an electronic version).

To start with, four different kinds of precipitants were tried, ammonium sulphate, methanol, PEG400 and PEG6000 respectively. The starting point for ammonium sulphate was set at 50%, which gave precipitation immediately. In later trials the concentration was decreased to 25% and 10%. The complex precipitated at a condition of 25% ammonium sulphate after approximately one week whereas interestingly conditions with 10% ammonium sulphate remained clear for quite a long time (more than four weeks), suggesting that 15% to 20% ammonium sulphate can be tried further. It should be noted that the pH did not affect the protein behaviour in these conditions.

Similarly, the complex precipitated in a short time in 50% methanol as a beginning point. Then the concentration was again decreased to 25% and 10%. Being an organic solvent, 25% methanol still denatured the complex after one day. However, until the end of recording after four months, conditions with 10% methanol were still clear. These results indicate that the metastable zone provided by methanol is very narrow and more effort needs to be taken to optimize these conditions.

As for PEG and its derivatives, 40% PEG400 and 10% PEG6000 was used at the start of screening. Interestingly, there was a significant difference between acidic and alkaline conditions with 40% PEG400. In acidic buffer, PEG400 denatured the complex immediately after setting the plates, whereas in alkaline buffer it took approximately two weeks to precipitate. Thus the amount of precipitant in acidic buffer was decreased from 40% to 20% and 10%, and raised the amount in alkaline buffer to 60% and 80%. It showed that the acidic conditions with PEG400 were dangerous to the complex, because it precipitated even at low PEG400 concentration. It is worth mentioning that 60% and 80% PEG400 in alkaline buffer denatured protein in a short time, showing that concentrations ranging from 40% to 60% can be tested further. The amount of PEG6000 was decreased to 5% and 2.5% because of precipitation at 10% PEG6000. The complex precipitated again at 5% PEG6000 regardless of the pH and a slow rate of precipitation was found at 2.5% PEG6000. It was observed that the complex was more stable in acidic conditions rather than alkaline conditions with 2.5% PEG6000, in contrast to those with PEG400. This suggests that as for PEG6000, the combination of pH, salt and precipitants

need screening further. (All the information has been summarized in sheets 1218 a, b and 1219 a–d, table 8.2)

In addition, we also set up crystals in-house according to the conditions obtained from HTS lab as described in the previous section. However, on the contrary to what happened from the high-throughput assays with sitting drop methods, proteins precipitated directly or several days later after hanging drop plates were set up (Information can be found in sheets 0125 a-g and 0222 a-f, table 8.2). In order to optimize the conditions, we increased the drop volume and decreased the concentration of precipitants. Despite the tree-like salt crystals, as observed in photographs from HTS lab, when this condition was reproduced, the complex still precipitated under most conditions (sheets 0306 a-d and sheets 0327 a-e, table 8.2). These results first confirm that crystals in figure 8.2 B, C & E are actually salt crystals. Besides, it also indicates that crystallization solution kits made in-house are, to some extent, different from those in HTS lab because of the precipitation we saw in-house, which should be taken into account in further studies.

8.4 Summary

In this Chapter efforts to crystallize the Vif/CBFβ/EloB/EloC complex are described and results from the trials are concluded. The high-throughput experiments in the HTS lab, HWI, did not provide as encouraging results as we had expected, suggesting crystallizing the complex is still one of the difficult problems in viral protein structural studies. The results from screening conditions in-house that were similar to those obtained from the HTS lab were discussed, i.e. salt crystals were observed in some conditions. Although the tetramer was soluble in its final buffer for over one week, when mixing with crystallization solution, for the majority of the conditions, this leaded to protein precipitation in a short time. This means that there is still a solubility problem for this complex. As a consequence, crystallization of the tetramer was unsuccessful, and the structure of the complex was not pursued further.

CHAPTER 9: DISCUSSION

9.1 SOCS-Elongin BC interaction

9.1.1 Comparison of the structures of the SOCS-box family

The SOCS-box family (SOCS1-7) proteins are indispensable regulators, functioning in many aspects including ubiquitination and transcription, and having high sequence similarity. Published structures show that they all share a common α-helical structure. It is worth noting, however, that the downstream sequences have a different spatial portfolio depending on the structural family. Domains from SOCS2 and SOCS4 (PDB ID: 2C9W, 2IZV) contains three small α-helixes that are known to interact with the EloB C-terminus as well (figure 9.1A, B) (Bullock, Debreczeni *et al*. 2006, Bullock, Rodriguez *et al*. 2007), whereas SOCS6 (PDB ID: 2VIF) adopts a partially folded sequence (figure 9.1C) and the SOCS3 (PDB ID: 3DCG) downstream sequence cannot be resolved in the X-ray crystallographical analysis (figure 9.1D) (Stanley, Ehrlich *et al*. 2008, Zadjali, Pike *et al*. 2011), which implies that the downstream structure is flexible in the solution and can be only observed by NMR. Our results for the first time report the whole SOCS3 domain including a BC-Box and the proline-rich motif in the presence of EloBC and show its conformational changes when interacting with EloBC.

9.1.2 Does the proline-rich motif perform other functions?

It has been demonstrated in our previous published work (Bergeron, Huthoff *et al*. 2010) and in Chapter 6 that mutations in the PPLPS motif impact the binding of Vif to the ubiquitination complex as assessed using co-IP assay. In theory, it is not rigorous to conclude from these kind of experiments, namely co-IP, that some residues on component A are crucial for the binding to component B because protein A and B may connect to other cellular factors that form a larger complex in cells, i.e. they do not bind directly to each other, contrary to

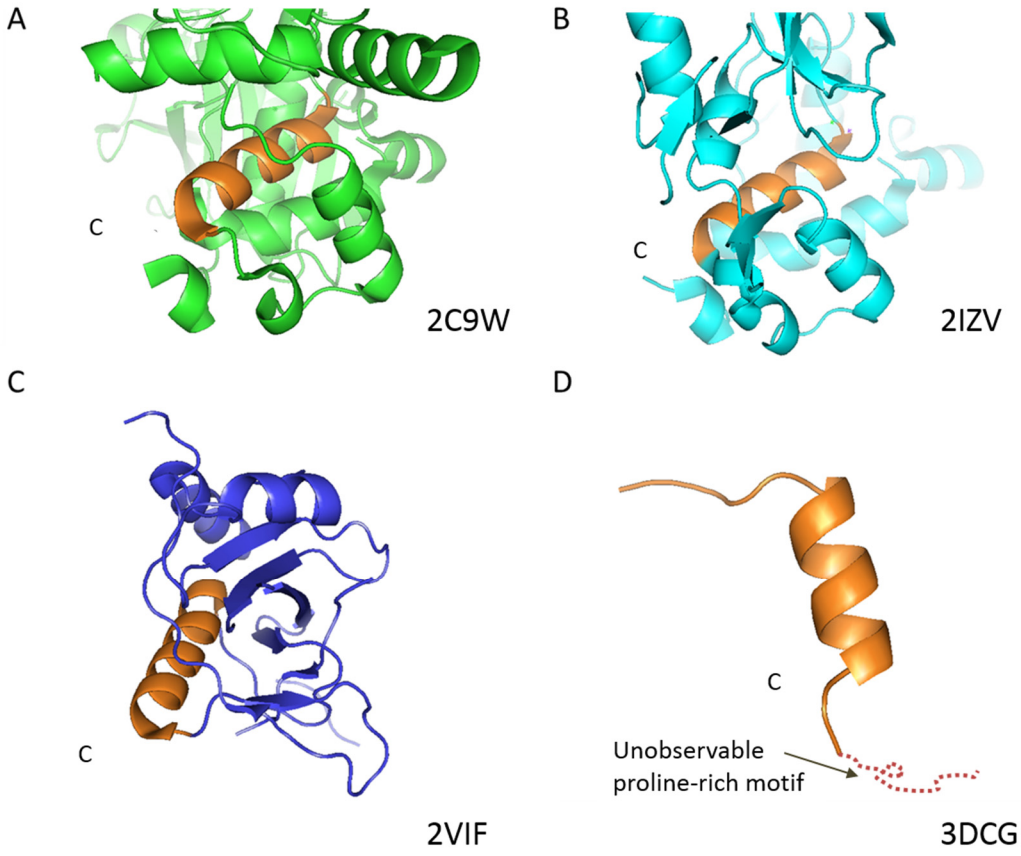


Figure 9.1: Structures of SOCS-box family proteins. Close up of the SOCS domains from published structures of different SOCS-box family. The C-terminus of each SOCS α-helical domain is labeled. (A) SOCS2 domain (2C9W). (B) SOCS4 domain (2IZV). (C) SOCS6 domain (2VIF). (D) SOCS3 domain (3DCG). The unobserved downstream sequence of SOCS3 is presented by a dotted line. The SOCS-domain on each structure is shown in brown.

what we observe in experiments. In this case, since it was proved that the PPLPS motif is not critical for the binding to EloBC, it can be inferred that this motif interacts with Cul5 although it is possible that there is an unknown factor existing in this complex. Therefore, further experiments are required to investigate this.

In previous studies it has been described that in the absence of SOCS-box the carboxyl terminus of EloB is very flexible in solution, and that the binding to SOCS-box induces the disordered DVMK stretch to gain structure. Considering the weak binding between these two domains which is not essential for SOCS-EloBC binding, we further suggest that it is the α-helix of EloC that first drives the induced-folding binding of Vif, followed by the interaction between the PPLPS motif and EloB C-terminal tail. The final common interface formed by the PPLPS motif and the induced-folding DVMK region in space is required to form an E3 ubiquitin ligase complex, properly specifically to recruit Cul5 (figure 9.2).

Some research studies report that the PPLPS motif also participates in forming Vif homomultimers (Yang, Sun *et al*. 2001, Yang, Gao *et al*. 2003), arguing that Vif multimerization is required for Vif function. Although Vif multimerization was observed both by *in vivo* or *in vitro* experiments, evidence in those studies, however, is not sufficient to prove that the multimerization of Vif is crucial in Vif-induced degradation on APOBEC3G. First, it is shown that the HCCH zinc-finger-like motif and BC-box play an important role in the pathway. From a structural biological viewpoint, the multimerization driven by the PPLPS motif may hide important functional domains. Moreover, in early studies (Yang, Sun *et al*. 2001) a ^{35}S-lableled Vif generated from *in vitro* translation was utilized for binding assays. It is doubtful that this produced Vif folded properly and that it kept its natural function, because the problems of expressing and purifying full-length Vif are still not solved until now. Techtmann and co-workers' work in 2012 showed that the PPLPS motif was not essential for Vif oligomerization and that the multimerization mechanism was more complicated and involved the HCCH domain, BC-box and downstream residues of Vif (Techtmann, Ghirlando *et al*. 2012). Therefore, a role of the PPLPS motif in Vif multimerization is plausible.

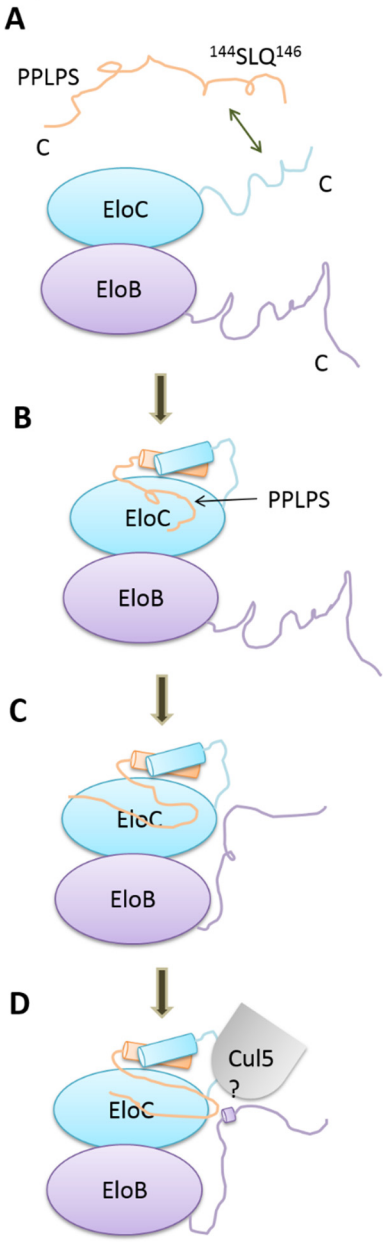


Figure 9.2: Schematic of the proposed induced-folding mechanism. The SOCS-box includes an α-helix domain (represented as ^{144}SLQ^{146}) and a proline-rich domain (represented as PPLPS). The formation of the Vif SOCS-EloBC complex is mainly driven by hydrophobic interactions between Vif BC-box and EloC C-terminus via a conformational change process. Then the proline-rich motif induces the EloB C-terminal tail folded to form a common interface to recruit cellular factors, perhaps Cul5. The various interaction events are presented in succession. The C-terminus of each subunit is indicated.

9.1.3 A dynamic model of SOCS-Elongin BC-Cullin5 interaction

It is commonly known that Vif hijacks APOBEC3G and recruits cellular factors Rbx, Cul5, EloB, EloC and CBFβ to stimulate the ubiquitination pathway on APOBEC3G. Figure 9.3 summarizes current knowledge on the complex. There is no report on the interaction between Rbx and Vif, and it has been shown in our studies that the PPLPS motif is not important for SOCS-EloBC binding. Many studies have confirmed that a zinc-finger-like motif, $H^{108}x_5Cx_{17-18}Cx_5H^{139}$, binds to Cul5 even in the absence of EloC (Luo, Xiao *et al*. 2005, Mehle, Thomas *et al*. 2006, Paul, Cui *et al*. 2006). Then a question arose why mutations in 144SLQ146 or in the PPLPS motif are also able to eliminate the ability of Vif to bind to the ubiquitination complex?

There is no doubt the static graphic picture displayed in figure 9.3 fails to answer this question. In order to explain all the results obtained from the experiments clearly, a dynamic analysis must be taken, i.e. the final Vif-APOBEC3G-cellular factor complex that is supposed to be true is assembled in a strict and concerted fashion. Any interference may stop the ubiquitination pathway. Therefore, here we further propose the following mechanism of the formation of this E3 ligase. As shown in figure 9.4, Vif is firstly translated and folds with the help of CBFβ in cells. After Vif detects APOBEC3G, it binds to the EloBC heterodimer by a hydrophobic interaction between Vif BC-box and EloC carboxyl domain. Then the PPLPS motif induces the folding of the EloB DVMK stretch at the C-terminal tail, forming a common interface. This Vif-EloBC interaction perhaps enables Vif to change the conformation of its HCCH and PPLP domain to acquire the successful positioning. Then these two domains, along with EloC, function together to recruit Cul5, stimulating the degradation on APOBEC3G. In this step, the two sites on Vif except for EloC are required to stabilize the Vif-ligase complex. This hypothesis explains why mutations in the zinc-finger-like motif, BC-box or the proline-rich motif respectively can disable the binding of Vif to Cul5 by pull-down assays or co–IP assays.

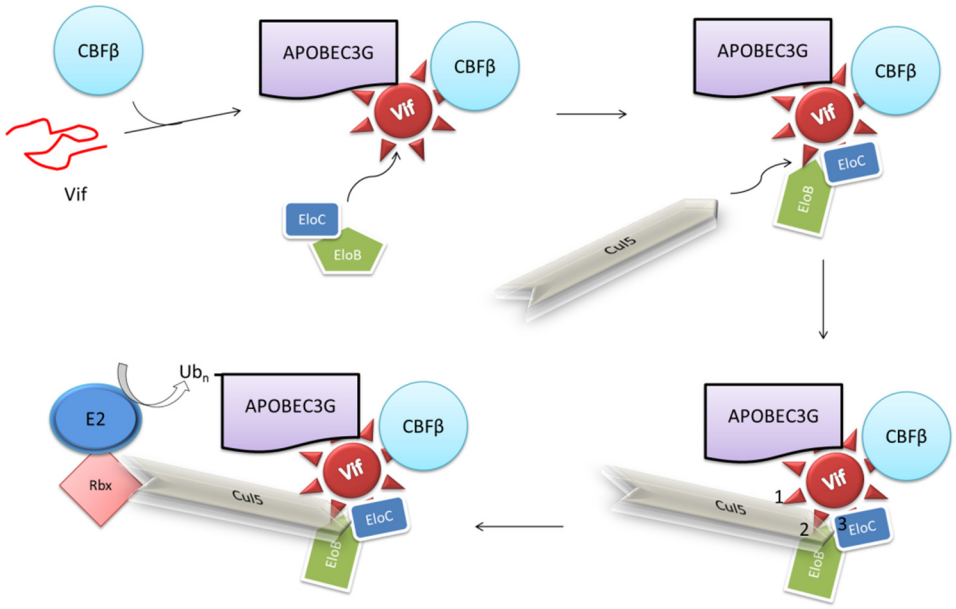


Figure 9.3: Vif forms an E3 complex to promote the degradation on A3G in a strict and concerted fashion. CBFβ helps Vif fold first in cells, followed by recognizing APOBEC3G. The Vif-A3G complex recruits EloBC and then Cul5. The formed ligation complex stimulates the degradation of APOBEC3G.

9.2 Problems within NMR studies on SOCS-Elongin BC complex

9.2.1 Unobserved peaks resulting from binding

Although the NMR solution structure of the complex was successfully solved by alternative NMR techniques, the occurrence of missing peaks from the spectra of the proteins of interest is still worth mentioning. As has been illustrated in previous Chapters, this problem is the main issue that hindered the progress of this project. Interestingly, the reason for the inability to observe peaks from SOCS-box and EloBC are different.

As for the SOCS-box, peaks from BC-box are observable in the free state, i.e. in the monomeric state. During the binding to EloBC, these peaks started disappearing rather than shifting or re-appearing at other positions. This is because of the exchange between two or more conformations in different micro-environments, which has a profound effect on the appearance and the chemical shifts of a resonance spin system. Basically, in two conditions, a spin may experience a conformation change. If the relaxation properties of each conformation are various under different conditions, the spectral properties will change. Figure 9.5 shows a brief example of exchange in two chemical environments. In conformation II, the methyl group may experience a ring current shift due to its proximity to the aromatic ring. Since the exchange phenomenon usually involves a change of the chemical environment of a spin, it is generally referred to as chemical exchange.

Mathematically, the exchange rate is defined and derived from a series of parameters involving rate constants of a chemical reaction and frequencies of spin states. Consider an exchange reaction given by the following scheme:

$$\mathbf{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{B}$$

where k_1 and k_2 are the kinetic rate constants for the conversion of A to B and B to A respectively, giving an overall equilibrium constant $K_{eq} = k_1/k_2$. Then the fraction of each conformation found in the system is:

$$p_A = \frac{k_2}{k_1 + k_2} = \frac{1}{1 + K_{eq}} \tag{9.1}$$

$$p_B = \frac{k_1}{k_1 + k_2} = \frac{K_{eq}}{1 + K_{eq}} \tag{9.2}$$

Now we define an apparent exchange rate, $k_{ex} = k_1 + k_2$ and a frequency difference between two conformational states, $\Delta\omega = \omega_A - \omega_B$. Basically, at a fast exchange rate ($k_{ex} >> \Delta\omega$), in the midway of a spin precessing, the conformation of the molecule is detected and changed to environment B. Consequently the precessional frequency changes to ω_B and meanwhile the rotation changes to the opposite direction. Finally the magnetization is along the original orientation. On the other hand, one slow exchange rate will give a spin sufficient time to precess irrespectively of a spin state before the molecular conformation changes. At the end signals from these two states are able to be acquired respectively, giving two peaks in the spectrum.

Besides, there is a simple way to distinguish fast from slow exchange. We define a parameter that represents the contribution of exchange to the overall spin-spin relaxation rate:

$$R_{ex} = \frac{p_A p_B k_{ex}}{1 + (k_{ex} / \Delta\omega)^2} \tag{9.3}$$

Here, another dimensionless parameter is used to characterize the time scale of chemical exchange has been defined by Palmer *et al* (Oscar Millet, J. Patrick Loria *et al*. 2000):

$$\alpha = \frac{d \ln R_{ex}}{d \ln \Delta\omega} \tag{9.4}$$

Substituting R_{ex} in equations 9.4 with 9.3, it can be derived that:

$$\alpha = \frac{2 / (k_{ex} / \Delta\omega)^2}{1 + (k_{ex} / \Delta\omega)^2} \tag{9.5}$$

Therefore it is clear that α varies from 0 for very slow exchange ($k_{ex} << \Delta\omega$) to 2 for very fast exchange ($k_{ex} >> \Delta\omega$). To summarize, in the fast exchange condition, the exchange process is capable of averaging the chemical shifts when the spins are precessing. Table 9.1 and figure 9.6 conclude the relationship between exchange rate and lineshape.

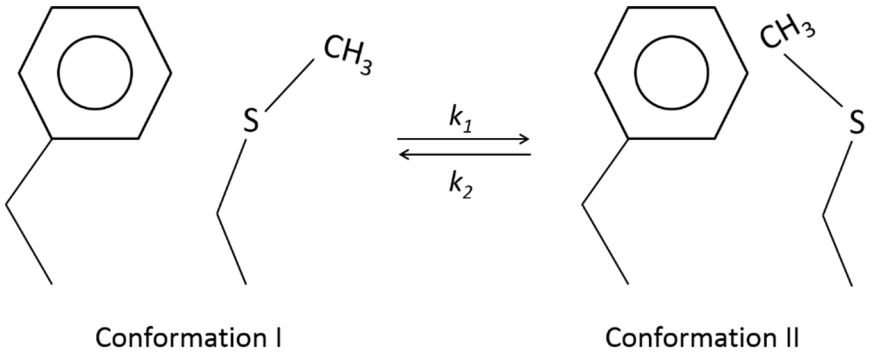


Figure 9.4: An example of conformational exchange.

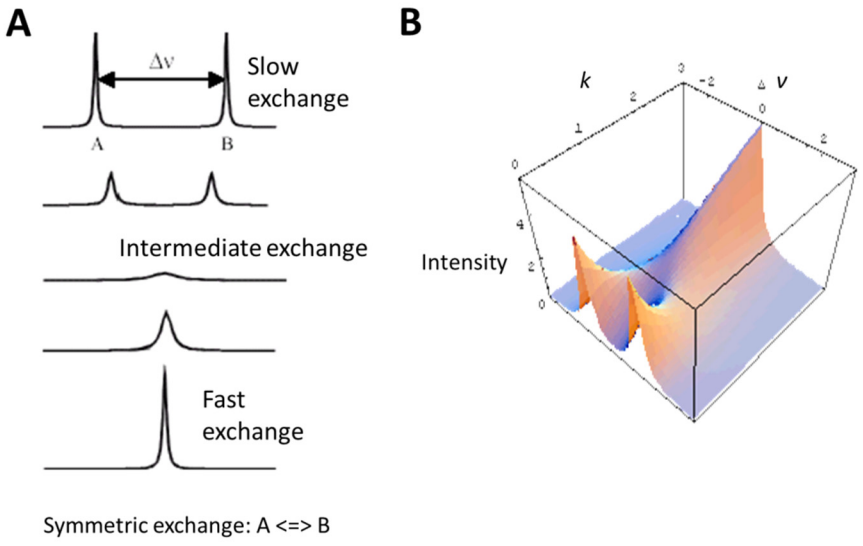


Figure 9.5: The effect of chemical exchange on lineshape. (A) The peak height changes alone with conformational exchange (http://triton.iqfr.csic.es/guide/eNMR/restraints/exchange.html) (B) Computational simulation of lineshapes: (http://www.bionmr.com/forum/nmr-software-15/programs-simulation-lineshape-change-due-chemical-exchange-140/).

Table 9.1, Summary of the effects of exchange on NMR spectra

Exchange rate	$k_{ex}/\Delta\omega$	**α**	**Observed spectrum**
Very slow	$k_{ex} << \Delta\omega$	0	Two resonances
Slow	$k_{ex} < \Delta\omega$	<1	Two broadened resonances
Intermediate	$k_{ex} \approx \Delta\omega$	1	Broad lineshape
Fast	$k_{ex} > \Delta\omega$	>1	One broadened resonance
Very fast	$k_{ex} >> \Delta\omega$	2	One sharp resonance

As for the EloBC dimer, besides the exchange rate, a slow correlation time due to a high molecular weight is the other reason that causes the disappearance of peaks. As was described in Chapter 2, NMR has two relaxation rates, T_1 and T_2. During T_2 relaxation, the proton T_2 of large proteins is:

$$\frac{1}{T_2} \approx \frac{9d^2}{20}\tau_c \tag{9.6}$$

and the nitrogen T_2 is:

$$\frac{1}{T_2} \approx \frac{d^2}{20}4\tau_c + \frac{1}{45}\omega_N^2\Delta\sigma^2 4\tau_c \tag{9.7}$$

where $d^2 = \gamma_H^2\gamma_N^2\hbar^2 / r^6$ and r is the inter-proton distance. Δσ represents the anisotropic chemical shift. Because large proteins tumble slower in solution, the rotational correlation time τ_c is mathematically proportional to the molecular weight. Thus for T_2, according to equations 9.6 and 9.7, it is inversely proportional to the molecular weight. Therefore, for large proteins the magnetization between spins relaxes much faster, resulting in less time to fully detect the signals. This in turn produces broader or weaker peaks, and ultimately lower signal-to-noise in NMR spectra. That is why in an EloBC spectrum only approximately 60% of the amino acids of a 216-residue protein can be observed.

9.2.2 The problem of Elongin B and Elongin C solubility

Since the complex of interest is a hetero-trimer, ideally it would be better to label each component independently, which may provide clearer spectra and allow assignment and ability to follow peak changes on binding. Therefore attempts to achieve preparations of soluble EloB and EloC were made. Expression at low temperature, in minimal media or low IPTG induction still failed to permit extraction of EloB or EloC from the inclusion bodies. Considering that the peaks from SOCS-box were clear, we tried to co-express SOCS-EloB and SOCS-EloC, thus extracting EloB and EloC peaks from the combined spectra. This alternative also failed due to the solubility problem.

There are several factors that affect the expression of a soluble protein in prokaryotic systems, e.g. *E. coli*. First, *E. coli* lacks chaperones that are required for protein translation and folding in eukaryotic cells, especially in mammalian cells. In order to address this problem, some companies engineered *E. coli* to express a group of human chaperones within these prokaryotic cells. Besides, some proteins are toxic to *E. coli*, and hence suppress the growth of cells or the expression of the proteins. The third factor is the characteristics of a protein sequence. Until now there are no general ways to solve the later two problems in protein studies. One alternative is to engineer the protein sequences, thus improving the solubility and decreasing the impact on host cells. The work on APOBEC3G is a typical case suggesting that engineering a sequence can dramatically facilitate protein preparation and structural calculation (Chen, Harjes *et al*. 2008).

9.3 Structural studies on Vif

Vif is a small accessory protein in HIV-1 rather than a viral structural protein. However, in the past dozens of years there have been great advances in the knowledge of the virus structure and the process of cell fusion. Owing to the discovery of CBFβ, it is possible to express and purify Vif from prokaryotic systems from which the yield of protein is promising for further biological studies. Considering that the Vif sample precipitated in most of the crystallization

conditions, there are still several problems that hinder the progress of structural studies of Vif.

First, the behaviour of Vif in the solution is not as stable as it should be regardless of the presence of the detergent. The solution is only clear at 4 ℃ whereas at room temperature it becomes cloudy within a short time. The purified sample was re-loaded onto a sizing column after one week, and only a small amount of protein was found to form the large aggregations observed after elution, suggesting that it may be kept at lower temperature for approximately 10 days. In addition, co-expressing Vif, CBFβ, EloC with a truncated EloB, M1-98 (residue number) and M1-106 respectively was tried in order to crystallize the complex without the flexible C-terminal tail of EloB. Interestingly, it is noticed that in the complex with EloB M1-106, the rate of precipitation after purification was faster than that of the wild type, whereas in the presence of M1-98 the complex precipitated directly after the elution from a Ni-NTA column. This phenomenon indicated that a strong self-association exists among Vif molecules in solution although they can be purified successfully, which can be seen from the sample of early elution fractions of gel filtration column (figure 8.1). The flexible C-terminal tail of EloB actually improves the solubility of the complex and, to some extent, decreases the self-interaction caused by Vif, thus stabilizing the complex in solution.

Moreover, in the complex of Vif-CBFβ-EloB-EloC, there are two known flexible elements, the hexa-histidine tag at the N terminus of EloB and the flexible C-terminus of EloB. In crystallography, flexible elements in a protein may prevent the formation of crystals by disturbing the nucleation because of their flexibility and dynamic nature. This is the reason why in many studies the His-tag is cut off before crystallization. In our case, it is possible that the retained His_6-tag and the C-terminus of EloB affects the formation of crystals. Now that the C-terminus of EloB has been shown to be crucial for the stability of the complex, removing the His-tag may be worth trying.

Another point is that CHAPS detergent is required in current protocol. In crystallography, because of special characteristics of a detergent, the behaviour of proteins in a system and the saturation of buffer are changed. Therefore, high salt concentration (usually over 0.5 M) is sometimes used to replace detergents, based on the idea that a strong solution background by salts can force protein to fold and to be stable.

This problem exists in analytical ultracentrifugation (AUC) field as well. The analysis on

proteins with detergents is actually a complex field for AUC. We used to run an AUC on our protein in the final buffer in order to confirm the molecular weight of the complex. However, the centrifugation data could not be used to calculate the molecular weight because the detergent that bound to the protein changed the hydrophobicity and molecular mass. It indicates that there may be more than one form of protein plus detergent existing in the solution, which may impair crystallization.

In summary, crystallization will only occur under very specific conditions. Any trace of ions and slight differences of protein concentrations, detergents or pH will impact on the characteristics of solution, especially free energy of solution and water activity. The problems described above prevent us from preparing crystals. More buffer condition optimizations are needed in the future.

9.4 Future work on HIV-1 Vif and relative factors

Since the discovery of CBFβ, the characteristics of Vif-CBFβ interaction and the binding regions in the complex of CBFβ-Vif-Cul5-EloBC have been reported (Zhou, Evans *et al*. 2012, Du, Zhao *et al*. 2013, Kim, Kwon *et al*. 2013), structural biologists and virologists have more understanding on how CBFβ helps Vif fold and keep stable and soluble *in vitro*. However, regardless of the huge progress in this field, there are no further studies reported on full-length Vif structure until now, suggesting that it is still difficult to crystallize and solve the structure of Vif. Combining our experience in the studies, there are other alternatives worth trying to crystallize Vif in the presence of some cellular factors.

As discussed above, the His-tag can be removed to produce an untagged tetramer. In order to achieve this, a protease cleavage site should be engineered between the His-tag and EloB. Before proteins of interests are eluted from a Ni-NTA column, a protease aliquot is loaded on to the column, thus eluting target proteins automatically. It prevents using high-concentration imidazole. Imidazole is known to be toxic to proteins in some cases, e.g. decreasing the activity (Lu, Chen *et al*. 2010). It is unclear whether high-concentration imidazole impacts the stability of the tetramer. In addition, in a typical affinity chromatography protocol, target proteins are

usually eluted with high-concentration imidazole (more than 200 mM), which may elute strongly bind impurities along with the proteins. Therefore, this method could improve the purity after a Ni-NTA column purification, thus raising the efficiency of gel filtration separation.

Another way is to co-crystallize the tetramer with more cellular factors, namely Cul5 or A3G. It has been reported that by small-angle X-ray scattering the complex of CBFβ-Vif-Cul5-EloBC is stable in solution with the molecular size that agrees with the theoretical value (Kim, Kwon *et al*. 2013). In experience, a protein with solubility problem can be very hard to be crystallized because it is easily precipitated out of the solution in crystallization kits in the presence of precipitants. Cul5, CBFβ and EloBC are thought to increase the solubility of Vif by sheltering the hydrophobic residues on the surface of Vif, which means that the large complex may stack to form crystals more easily than the Vif-EloBC-CBFβ complex. As for A3G, there is no report on the structure of N-terminus because of its insolubility. If this binding has a high affinity, it indicates that the generated product may be stable enough for crystallization. In order to achieve this, N-terminal A3G can be co-expressed with Vif and CBFβ, followed by pull-down assay with purified EloBC on a Ni-NTA column. This way would provide more interesting information on the Vif-APOBEC3G interaction.

In a recent report Kim's experiments demonstrated that Vif and CBFβ bind to each other mainly by the N-terminal 140 residues (Kim, Kwon *et al*. 2013). Our work has solved the structure of downstream SOCS-box. Therefore, it is also worth trying to study the small dimer, i.e. Vif140-CBFβ140, instead of an entire large complex by structural biological methods. It is of note that the molecular weight of this 280-residue dimer is around 30 kDa, which is still suitable for NMR studies. Encouragingly, now that the assignment of N-terminal region of CBFβ (residues 1-138) and the NMR solution structure has been published (Goger, Gupta *et al*. 1999), it would dramatically facilitate the NMR work on the dimer. In this case, the advantage of NMR that it does not require crystals can be used to identify the structure of the dimer and even the dynamics among Vif-CBFβ-Cul5. By using this approach maybe a Vif structure can be elucidated.

9.5 Concluding remarks

The work have strengthened the current knowledge on the interaction of Vif SOCS-box and EloBC heterodimer by calculating the structure of the trimer and studying the specific motifs or elements. The results obtained in this work and our previous studies (Bergeron, Huthoff *et al*. 2010) have clearly uncovered that SOCS-box, EloB and EloC all experience structural changes, and that the binding between SOCS-EloC and SOCS-EloB happens successively. Now that a significant biological meaning of the SOCS-EloB interaction was not observed in our biophysical and biochemical experiments, an important question emerges that what the particular role of this interaction is in cells. We deduce that the common interface is formed to bind Cul5. Regarding that EloBC and Cul5 are endogenous and crucial for cell survival, in the next step it is worth testing the interaction between the purified trimer and purified Cul5 using biophysical and biochemical assays *in vitro*.

Another interesting aspect is the comparison among different SOCS-box-containing proteins. As has been discussed above, the structures of SOCS family can be classified into two groups, namely a completely-structured group and a partially-structured group. However, the biological meanings of the structural differences are still unclear. In the future a systematic comparison may be required to investigate the functions of different regions, which will provide a more comprehensive understanding of Vif SOCS-box.

Finally in this work we present a protocol of expression and purification of high-yield folded and soluble Vif from *E. coli* in the presence of EloBC and CBFβ, which facilitates the structural studies of Vif and its binding partners. Efforts have been made to crystallize the tetramer whereas there were no positive results from HTS assays and in-house screening. In the future more conditions need exploring and some new constructs mentioned above are worth trying due to their individual advantages. In addition, our studies also suggest that the C-terminus of CBFβ is important for Vif folding as well as the N-terminal 140 resides (Kim, Kwon *et al*. 2013). The dynamics mechanism of CBFβ-induced Vif folding is another important topic that can be further investigated from a structural biological standpoint.

REFERENCES :

Aiken, C., J. Konner, N. R. Landau, M. E. Lenburg and D. Trono (1994). "Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain." Cell **76**(5): 853-864.

Albin, J. S., J. S. Anderson, J. R. Johnson, E. Harjes, H. Matsuo, N. J. Krogan and R. S. Harris (2013). "Dispersed sites of HIV Vif-dependent polyubiquitination in the DNA deaminase APOBEC3F." J Mol Biol **425**(7): 1172-1182.

Albin, J. S. and R. S. Harris (2010). "Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics." Expert Rev Mol Med **12**: e4.

Anderson, W. A. and R. Freeman (1962). "Influence of a second radiofrequency field on high-resolution nuclear magnetic resonance spectra." J Chem Phys **37**(1): 5.

Archin, N. M., A. Espeseth, D. Parker, M. Cheema, D. Hazuda and D. M. Margolis (2009). "Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid." AIDS Res Hum Retroviruses **25**(2): 207-212.

Archin, N. M., K. S. Keedy, A. Espeseth, H. Dang, D. J. Hazuda and D. M. Margolis (2009). "Expression of latent human immunodeficiency type 1 is induced by novel and selective histone deacetylase inhibitors." AIDS **23**(14): 1799-1806.

Archin, N. M., A. L. Liberty, A. D. Kashuba, S. K. Choudhary, J. D. Kuruc, A. M. Crooks, D. C. Parker, E. M. Anderson, M. F. Kearney, M. C. Strain, D. D. Richman, M. G. Hudgens, R. J. Bosch, J. M. Coffin, J. J. Eron, D. J. Hazuda and D. M. Margolis (2012). "Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy." Nature **487**(7408): 482-485.

Arhel, N. J. and F. Kirchhoff (2009). "Implications of Nef: host cell interactions in viral persistence and progression to AIDS." Curr Top Microbiol Immunol **339**: 147-175.

Aso, T., W. S. Lane, J. W. Conaway and R. C. Conaway (1995). "Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II." Science **269**(5229): 1439-1443.

Ayyavoo, V., K. Muthumani, S. Kudchodkar, D. Zhang, P. Ramanathan, N. S. Dayes, J. J. Kim, J. I. Sin, L. J. Montaner and D. B. Weiner (2002). "HIV-1 viral protein R compromises cellular immune function in vivo." Int Immunol **14**(1): 13-22.

Babon, J. J., J. K. Sabo, A. Soetopo, S. Yao, M. F. Bailey, J. G. Zhang, N. A. Nicola and R. S. Norton (2008).

"The SOCS box domain of SOCS3: structure and interaction with the elonginBC-cullin5 ubiquitin ligase." J Mol Biol **381**(4): 928-940.

Balci, M. (2005). Basic ^{1}H- and ^{13}C-NMR spectroscopy. Amsterdam ; Boston, Elsevier.

Banks, K. M., D. R. Hare and B. R. Reid (1989). "Three-dimensional solution structure of a DNA duplex containing the BclI restriction sequence: two-dimensional NMR studies, distance geometry calculations, and refinement by back-calculation of the NOESY spectrum." Biochemistry **28**(17): 6996-7010.

Bannwarth, S. and A. Gatignol (2005). "HIV-1 TAR RNA: the target of molecular interactions between the virus and its host." Curr HIV Res **3**(1): 61-71.

Barraud, P., J. C. Paillart, R. Marquet and C. Tisne (2008). "Advances in the structural understanding of Vif proteins." Curr HIV Res **6**(2): 91-99.

Battiste, J. L. and G. Wagner (2000). "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data." Biochemistry **39**(18): 5355-5365.

Bax, A., G. M. Clore and G. A. M. (1990). "1H-1H Correlation Via Isotropic Mixing of 13C Magnetization, a New 3-Dimensional Approach for Assigning 1H and 13C Spectra of 13C-Enriched Proteins." J Magn Reson **88**(2): 7.

Bax, A. and M. Ikura (1991). "An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins." J Biomol NMR **1**(1): 99-104.

Bayer, P., M. Kraft, A. Ejchart, M. Westendorp, R. Frank and P. Rosch (1995). "Structural studies of HIV-1 Tat protein." J Mol Biol **247**(4): 529-535.

Benko, Z., D. Liang, E. Agbottah, J. Hou, L. Taricani, P. G. Young, M. Bukrinsky and R. Y. Zhao (2007). "Antagonistic interaction of HIV-1 Vpr with Hsf-mediated cellular heat shock response and Hsp16 in fission yeast (Schizosaccharomyces pombe)." Retrovirology **4**: 16.

Bergeron, J. R., H. Huthoff, D. A. Veselkov, R. L. Beavil, P. J. Simpson, S. J. Matthews, M. H. Malim and M. R. Sanderson (2010). "The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex." PLoS Pathog **6**(6): e1000925.

Berjanskii, M. V. and D. S. Wishart (2005). "A simple method to predict protein flexibility using secondary chemical shifts." J Am Chem Soc **127**(43): 14970-14971.

Berkhout, B., A. T. Das and N. Beerens (2001). "HIV-1 RNA editing, hypermutation, and error-prone

reverse transcription." Science **292**(5514): 7.

Berkhout, B. and A. de Ronde (2004). "APOBEC3G versus reverse transcriptase in the generation of HIV-1 drug-resistance mutations." AIDS **18**(13): 1861-1863.

Berkhout, B. and F. J. van Hemert (1994). "The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins." Nucleic Acids Res **22**(9): 1705-1711.

Bishop, K. N., M. Verma, E. Y. Kim, S. M. Wolinsky and M. H. Malim (2008). "APOBEC3G inhibits elongation of HIV-1 reverse transcripts." PLoS Pathog **4**(12): e1000231.

Bjelic, S. and I. Jelesarov (2008). "A survey of the year 2007 literature on applications of isothermal titration calorimetry." J Mol Recognit **21**(5): 289-312.

Blagoveshchenskaya, A. D., L. Thomas, S. F. Feliciangeli, C. H. Hung and G. Thomas (2002). "HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway." Cell **111**(6): 853-866.

Bloembergen, N., E. M. Purcell and R. V. Pound (1948). "Relaxation Effects in Nuclear Magnetic Resonance Absorption." Physical Review **73**(7): 34.

Bour, S., U. Schubert and K. Strebel (1995). "The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation." J Virol **69**(3): 1510-1520.

Brigati, C., M. Giacca, D. M. Noonan and A. Albini (2003). "HIV Tat, its TARgets and the control of viral gene expression." FEMS Microbiol Lett **220**(1): 57-65.

Briggs, J. A., K. Grunewald, B. Glass, F. Forster, H. G. Krausslich and S. D. Fuller (2006). "The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions." Structure **14**(1): 15-20.

Briggs, J. A., M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt and M. C. Johnson (2004). "The stoichiometry of Gag protein in HIV-1." Nat Struct Mol Biol **11**(7): 672-675.

Browne, E. P., C. Allers and N. R. Landau (2009). "Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent." Virology **387**(2): 313-321.

Bullock, A. N., J. E. Debreczeni, A. M. Edwards, M. Sundstrom and S. Knapp (2006). "Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase." Proc Natl Acad Sci U S A **103**(20): 7637-7642.

Bullock, A. N., M. C. Rodriguez, J. E. Debreczeni, Z. Songyang and S. Knapp (2007). "Structure of the SOCS4-ElonginB/C complex reveals a distinct SOCS box interface and the molecular basis for SOCS-

dependent EGFR degradation." Structure **15**(11): 1493-1504.

Casartelli, N., F. Guivel-Benhassine, R. Bouziat, S. Brandler, O. Schwartz and A. Moris (2010). "The antiviral factor APOBEC3G improves CTL recognition of cultured HIV-infected T cells." J Exp Med **207**(1): 39-49.

Cavanagh, J. (2007). Protein NMR spectroscopy : principles and practice. Amsterdam ; Boston, Academic Press.

Chan, D. C. and P. S. Kim (1998). "HIV entry and its inhibition." Cell **93**(5): 681-684.

Chen, J. C., J. Krucinski, L. J. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt and R. M. Stroud (2000). "Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding." Proc Natl Acad Sci U S A **97**(15): 8233-8238.

Chen, K. M., E. Harjes, P. J. Gross, A. Fahmy, Y. Lu, K. Shindo, R. S. Harris and H. Matsuo (2008). "Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G." Nature **452**(7183): 116-119.

Cochrane, A. W., A. Perkins and C. A. Rosen (1990). "Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function." J Virol **64**(2): 881-885.

Conticello, S. G., R. S. Harris and M. S. Neuberger (2003). "The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G." Curr Biol **13**(22): 2009-2013.

Conticello, S. G., C. J. Thomas, S. K. Petersen-Mahrt and M. S. Neuberger (2005). "Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases." Mol Biol Evol **22**(2): 367-377.

Contreras, X., M. Schweneker, C. S. Chen, J. M. McCune, S. G. Deeks, J. Martin and B. M. Peterlin (2009). "Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells." J Biol Chem **284**(11): 6782-6789.

Control, C. f. D. (1982). "Update on acquired immune deficiency syndrome (AIDS) - United States." MMWR Morb Mortal Wkly Rep **31**(37): 4.

Dang, Y., X. Wang, W. J. Esselman and Y. H. Zheng (2006). "Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family." J Virol **80**(21): 10522-10533.

Daugherty, M. D., B. Liu and A. D. Frankel (2010). "Structural basis for cooperative RNA binding and export complex assembly by HIV Rev." Nat Struct Mol Biol **17**(11): 1337-1342.

de Vries, S. J., A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar and A. M. Bonvin (2007). "HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the

CAPRI targets." Proteins **69**(4): 726-733.

de Vries, S. J., M. van Dijk and A. M. Bonvin (2010). "The HADDOCK web server for data-driven biomolecular docking." Nat Protoc **5**(5): 883-897.

Deeks, S. G. (2012). "HIV: Shock and kill." Nature **487**(7408): 439-440.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

Delarue, M. (2008). "Dealing with structural variability in molecular replacement and crystallographic refinement through normal-mode analysis." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 40-48.

Deloche, B. and E. T. Samulski (1981). "Short-range nematic-like orientational order in strained elastomers." Macromolecules **14**(3): 7.

Deshaies, R. J. and C. A. Joazeiro (2009). "RING domain E3 ubiquitin ligases." Annu Rev Biochem **78**: 399-434.

DiMattia, M. A., N. R. Watts, S. J. Stahl, C. Rader, P. T. Wingfield, D. I. Stuart, A. C. Steven and J. M. Grimes (2010). "Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element." Proc Natl Acad Sci U S A **107**(13): 5810-5814.

Dimitrov, A. S., J. M. Louis, C. A. Bewley, G. M. Clore and R. Blumenthal (2005). "Conformational changes in HIV-1 gp41 in the course of HIV-1 envelope glycoprotein-mediated fusion and inactivation." Biochemistry **44**(37): 12471-12479.

Dominguez, C., R. Boelens and A. M. Bonvin (2003). "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information." J Am Chem Soc **125**(7): 1731-1737.

Donahue, J. P., M. L. Vetter, N. A. Mukhtar and R. T. D'Aquila (2008). "The HIV-1 Vif PPLPS motif is necessary for human APOBEC3G binding and degradation." Virology **377**(1): 49-53.

Doreleijers, J. F., A. W. Sousa da Silva, E. Krieger, S. B. Nabuurs, C. A. Spronk, T. J. Stevens, W. F. Vranken, G. Vriend and G. W. Vuister (2012). "CING: an integrated residue-based structure validation program suite." J Biomol NMR **54**(3): 267-283.

Douglas, J. L., Y. Bai, J. K. Gustin and A. V. Moses (2013). "A comparative mutational analysis of HIV-1 Vpu subtypes B and C for the identification of determinants required to counteract BST-2/Tetherin and enhance viral egress." Virology.

Douglas, J. L., K. Viswanathan, M. N. McCarroll, J. K. Gustin, K. Fruh and A. V. Moses (2009). "Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/Tetherin via a

{beta}TrCP-dependent mechanism." J Virol **83**(16): 7931-7947.

Du, J., K. Zhao, Y. Rui, P. Li, X. Zhou, W. Zhang and X. F. Yu (2013). "Differential requirements for HIV-1 Vif-mediated APOBEC3G degradation and RUNX1-mediated transcription by core binding factor beta." J Virol **87**(3): 1906-1911.

Dube, M., B. B. Roy, P. Guiot-Guillain, J. Binette, J. Mercier, A. Chiasson and E. A. Cohen (2010). "Antagonism of tetherin restriction of HIV-1 release by Vpu involves binding and sequestration of the restriction factor in a perinuclear compartment." PLoS Pathog **6**(4): e1000856.

Duda, D. M., L. A. Borg, D. C. Scott, H. W. Hunt, M. Hammel and B. A. Schulman (2008). "Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation." Cell **134**(6): 995-1006.

Dundr, M., G. H. Leno, M. L. Hammarskjold, D. Rekosh, C. Helga-Maria and M. O. Olson (1995). "The roles of nucleolar structure and function in the subcellular location of the HIV-1 Rev protein." J Cell Sci **108 (Pt 8)**: 2811-2823.

Eijkelenboom, A. P., R. Sprangers, K. Hard, R. A. Puras Lutzke, R. H. Plasterk, R. Boelens and R. Kaptein (1999). "Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase." Proteins **36**(4): 556-564.

Engelman, A. and P. Cherepanov (2012). "The structural biology of HIV-1: mechanistic and therapeutic insights." Nat Rev Microbiol **10**(4): 279-290.

Fang, J., S. Kubota, B. Yang, N. Zhou, H. Zhang, R. Godbout and R. J. Pomerantz (2004). "A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev." Virology **330**(2): 471-480.

Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo and F. Wong-Staal (1986). "HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA." Cell **46**(6): 807-817.

Feng, Y., C. C. Broder, P. E. Kennedy and E. A. Berger (1996). "HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272**(5263): 872-877.

Feng, Y., R. P. Love and L. Chelico (2013). "HIV-1 viral infectivity factor (Vif) alters processive single-stranded DNA scanning of the retroviral restriction factor APOBEC3G." J Biol Chem **288**(9): 6083-6094.

Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj and R. Luhrmann (1995). "The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs." Cell **82**(3): 475-483.

Fisher, A. G., B. Ensoli, L. Ivanoff, M. Chamberlain, S. Petteway, L. Ratner, R. C. Gallo and F. Wong-Staal (1987). "The sor gene of HIV-1 is required for efficient virus transmission in vitro." Science **237**(4817): 888-893.

Fogh, R. H., W. Boucher, W. F. Vranken, A. Pajon, T. J. Stevens, T. N. Bhat, J. Westbrook, J. M. Ionides and E. D. Laue (2005). "A framework for scientific data modeling and automated software development." Bioinformatics **21**(8): 1678-1684.

Foster, J. L. and J. V. Garcia (2008). "HIV-1 Nef: at the crossroads." Retrovirology **5**: 84.

Freed, E. O. (2001). "HIV-1 replication." Somat Cell Mol Genet **26**(1-6): 13-33.

Gandhi, S. K., J. D. Siliciano, J. R. Bailey, R. F. Siliciano and J. N. Blankson (2008). "Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors." J Virol **82**(6): 3125-3130.

Gatignol, A. and K. T. Jeang (2000). "Tat as a transcriptional activator and a potential therapeutic target for HIV-1." Adv Pharmacol **48**: 209-227.

Gautam, A. and J. Bhattacharya (2013). "Evidence that Vpu Modulates HIV-1 Gag-Envelope Interaction towards Envelope Incorporation and Infectivity in a Cell Type Dependent Manner." PLoS One **8**(4): e61388.

Gillim-Ross, L., A. Cara and M. E. Klotman (2005). "HIV-1 extrachromosomal 2-LTR circular DNA is long-lived in human macrophages." Viral Immunol **18**(1): 190-196.

Goffinet, C., I. Allespach, S. Homann, H. M. Tervo, A. Habermann, D. Rupp, L. Oberbremer, C. Kern, N. Tibroni, S. Welsch, J. Krijnse-Locker, G. Banting, H. G. Krausslich, O. T. Fackler and O. T. Keppler (2009). "HIV-1 antagonism of CD317 is species specific and involves Vpu-mediated proteasomal degradation of the restriction factor." Cell Host Microbe **5**(3): 285-297.

Goffinet, C., S. Homann, I. Ambiel, N. Tibroni, D. Rupp, O. T. Keppler and O. T. Fackler (2010). "Antagonism of CD317 restriction of human immunodeficiency virus type 1 (HIV-1) particle release and depletion of CD317 are separable activities of HIV-1 Vpu." J Virol **84**(8): 4089-4094.

Goger, M., V. Gupta, W. Y. Kim, K. Shigesada, Y. Ito and M. H. Werner (1999). "Molecular insights into PEBP2/CBF beta-SMMHC associated acute leukemia revealed from the structure of PEBP2/CBF beta." Nat Struct Biol **6**(7): 620-623.

Goh, W. C., N. Manel and M. Emerman (2004). "The human immunodeficiency virus Vpr protein binds Cdc25C: implications for G2 arrest." Virology **318**(1): 337-349.

Green, D. W., V. M. Ingram and M. F. Perutz (1954). "The Structure of Haemoglobin .4. Sign Determination by the Isomorphous Replacement Method." Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences **225**(1162): 287-307.

Gronenborn, A. M., D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield and G. M. Clore (1991). "A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G." Science **253**(5020): 657-661.

Grzesiek, S. and A. Bax (1993). "Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins." J Biomol NMR **3**(2): 185-204.

Guruprasad, K., B. V. Reddy and M. W. Pandit (1990). "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence." Protein Eng **4**(2): 155-161.

Hache, G., L. M. Mansky and R. S. Harris (2006). "Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance." AIDS Rev **8**(3): 148-157.

Hache, G., K. Shindo, J. S. Albin and R. S. Harris (2008). "Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G." Curr Biol **18**(11): 819-824.

Hare, S., S. S. Gupta, E. Valkov, A. Engelman and P. Cherepanov (2010). "Retroviral intasome assembly and inhibition of DNA strand transfer." Nature **464**(7286): 232-236.

Harris, R. S., K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger and M. H. Malim (2003). "DNA deamination mediates innate immunity to retroviral infection." Cell **113**(6): 803-809.

Harris, R. S., A. M. Sheehy, H. M. Craig, M. H. Malim and M. S. Neuberger (2003). "DNA deamination: not just a trigger for antibody diversification but also a mechanism for defense against retroviruses." Nat Immunol **4**(7): 641-643.

Haynes, B. F. and M. J. McElrath (2013). "Progress in HIV-1 vaccine development." Curr Opin HIV AIDS **8**(4): 326-332.

He, Z., W. Zhang, G. Chen, R. Xu and X. F. Yu (2008). "Characterization of conserved motifs in HIV-1 Vif required for APOBEC3G and APOBEC3F interaction." J Mol Biol **381**(4): 1000-1011.

Hetzer, C., W. Dormeyer, M. Schnolzer and M. Ott (2005). "Decoding Tat: the biology of HIV Tat posttranslational modifications." Microbes Infect **7**(13): 1364-1369.

Hochstrasser, M. (2009). "Origin and function of ubiquitin-like proteins." Nature **458**(7237): 422-429.

Hofmann, W., B. Reichart, A. Ewald, E. Muller, I. Schmitt, R. H. Stauber, F. Lottspeich, B. M. Jockusch, U. Scheer, J. Hauber and M. C. Dabauvalle (2001). "Cofactor requirements for nuclear export of Rev response element (RRE)- and constitutive transport element (CTE)-containing retroviral RNAs. An unexpected role for actin." J Cell Biol **152**(5): 895-910.

Holden, L. G., C. Prochnow, Y. P. Chang, R. Bransteitter, L. Chelico, U. Sen, R. C. Stevens, M. F. Goodman and X. S. Chen (2008). "Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications." Nature **456**(7218): 121-124.

Holmes, R. K., F. A. Koning, K. N. Bishop and M. H. Malim (2007). "APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G." J Biol Chem **282**(4): 2587-2595.

Homann, S., D. Smith, S. Little, D. Richman and J. Guatelli (2011). "Upregulation of BST-2/Tetherin by HIV infection in vivo." J Virol **85**(20): 10659-10668.

Hong, H. W., S. W. Lee and H. Myung (2013). "Induced degradation of Tat by nucleocapsid (NC) via the proteasome pathway and its effect on HIV transcription." Viruses **5**(4): 1143-1152.

Hope, T. J. (1999). "The ins and outs of HIV Rev." Arch Biochem Biophys **365**(2): 186-191.

Imahashi, M., M. Nakashima and Y. Iwatani (2012). "Antiviral Mechanism and Biochemical Basis of the Human APOBEC3 Family." Front Microbiol **3**: 250.

Iordanskiy, S., Y. Zhao, P. DiMarzio, I. Agostini, L. Dubrovsky and M. Bukrinsky (2004). "Heat-shock protein 70 exerts opposing effects on Vpr-dependent and Vpr-independent HIV-1 replication in macrophages." Blood **104**(6): 1867-1872.

Iordanskiy, S., Y. Zhao, L. Dubrovsky, T. Iordanskaya, M. Chen, D. Liang and M. Bukrinsky (2004). "Heat shock protein 70 protects cells from cell cycle arrest and apoptosis induced by human immunodeficiency virus type 1 viral protein R." J Virol **78**(18): 9697-9704.

Isel, C., J. M. Lanchy, S. F. Le Grice, C. Ehresmann, B. Ehresmann and R. Marquet (1996). "Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys." EMBO J **15**(4): 917-924.

Iwatani, Y., D. S. Chan, F. Wang, K. S. Maynard, W. Sugiura, A. M. Gronenborn, I. Rouzina, M. C. Williams, K. Musier-Forsyth and J. G. Levin (2007). "Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G." Nucleic Acids Res **35**(21): 7096-7108.

Jacobo-Molina, A., J. Ding, R. G. Nanni, A. D. Clark, Jr., X. Lu, C. Tantillo, R. L. Williams, G. Kamer, A. L.

Ferris, P. Clark and et al. (1993). "Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA." Proc Natl Acad Sci U S A **90**(13): 6320-6324.

Jager, S., D. Y. Kim, J. F. Hultquist, K. Shindo, R. S. LaRue, E. Kwon, M. Li, B. D. Anderson, L. Yen, D. Stanley, C. Mahon, J. Kane, K. Franks-Skiba, P. Cimermancic, A. Burlingame, A. Sali, C. S. Craik, R. S. Harris, J. D. Gross and N. J. Krogan (2012). "Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection." Nature **481**(7381): 371-375.

Janini, M., M. Rogers, D. R. Birx and F. E. McCutchan (2001). "Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells." J Virol **75**(17): 7973-7986.

Jarmuz, A., A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott and N. Navaratnam (2002). "An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22." Genomics **79**(3): 285-296.

Jern, P., R. A. Russell, V. K. Pathak and J. M. Coffin (2009). "Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance." PLoS Pathog **5**(4): e1000367.

Ji, H., W. Shu, F. T. Burling, S. Jiang and M. Lu (1999). "Inhibition of human immunodeficiency virus type 1 infectivity by the gp41 core: role of a conserved hydrophobic cavity in membrane fusion." J Virol **73**(10): 8578-8586.

Jia, B., R. Serra-Moreno, W. Neidermyer, A. Rahmberg, J. Mackey, I. B. Fofana, W. E. Johnson, S. Westmoreland and D. T. Evans (2009). "Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2." PLoS Pathog **5**(5): e1000429.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.

Joyner, J. C., K. D. Keuper and J. A. Cowan (2013). "Kinetics and Mechanisms of Oxidative Cleavage of HIV RRE RNA by Rev-Coupled Transition Metal Chelates." Chem Sci **4**(4): 1707-1718.

Kamura, T., K. Maenaka, S. Kotoshiba, M. Matsumoto, D. Kohda, R. C. Conaway, J. W. Conaway and K. I. Nakayama (2004). "VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases." Genes Dev **18**(24): 3055-3065.

Kao, S., R. Goila-Gaur, E. Miyagi, M. A. Khan, S. Opi, H. Takeuchi and K. Strebel (2007). "Production of

infectious virus and degradation of APOBEC3G are separable functional properties of human immunodeficiency virus type 1 Vif." Virology **369**(2): 329-339.

Kasper, M. R., J. F. Roeth, M. Williams, T. M. Filzen, R. I. Fleis and K. L. Collins (2005). "HIV-1 Nef disrupts antigen presentation early in the secretory pathway." J Biol Chem **280**(13): 12840-12848.

Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Greyson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn and G. M. Shaw (2008). "Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection." Proc Natl Acad Sci U S A **105**(21): 7552-7557.

Keeler, J. (2005). Understanding NMR spectroscopy. Chichester, England ; Hoboken, NJ, Wiley.

Kemp, W. (1987). Organic spectroscopy. Basingstoke, Macmillan.

Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-666.

Kieffer, T. L., P. Kwon, R. E. Nettles, Y. Han, S. C. Ray and R. F. Siliciano (2005). "G-->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo." J Virol **79**(3): 1975-1980.

Kim, D. Y., E. Kwon, P. D. Hartley, D. C. Crosby, S. Mann, N. J. Krogan and J. D. Gross (2013). "CBFbeta stabilizes HIV Vif to counteract APOBEC3 at the expense of RUNX1 target gene expression." Mol Cell **49**(4): 632-644.

Kirchhoff, F. (2010). "Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses." Cell Host Microbe **8**(1): 55-67.

Kirchhoff, F., T. C. Greenough, D. B. Brettler, J. L. Sullivan and R. C. Desrosiers (1995). "Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection." N Engl J Med **332**(4): 228-232.

Knauth, K., E. Cartwright, S. Freund, M. Bycroft and A. Buchberger (2009). "VHL mutations linked to type 2C von Hippel-Lindau disease cause extensive structural perturbations in pVHL." J Biol Chem **284**(16): 10514-10522.

Kobayashi, M., A. Takaori-Kondo, Y. Miyauchi, K. Iwai and T. Uchiyama (2005). "Ubiquitination of

APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function." J Biol Chem **280**(19): 18573-18578.

Kohli, R. M., R. W. Maul, A. F. Guminski, R. L. McClure, K. S. Gajula, H. Saribasak, M. A. McMahon, R. F. Siliciano, P. J. Gearhart and J. T. Stivers (2010). "Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification." J Biol Chem **285**(52): 40956-40964.

Kohlstaedt, L. A., J. Wang, J. M. Friedman, P. A. Rice and T. A. Steitz (1992). "Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor." Science **256**(5065): 1783-1790.

Korotkova, N., Y. Yang, I. Le Trong, E. Cota, B. Demeler, J. Marchant, W. E. Thomas, R. E. Stenkamp, S. L. Moseley and S. Matthews (2008). "Binding of Dr adhesins of Escherichia coli to carcinoembryonic antigen triggers receptor dissociation." Mol Microbiol **67**(2): 420-434.

Koszelak-Rosenblum, M., A. Krol, N. Mozumdar, K. Wunsch, A. Ferin, E. Cook, C. K. Veatch, R. Nagel, J. R. Luft, G. T. Detitta and M. G. Malkowski (2009). "Determination and application of empirically derived detergent phase boundaries to effectively crystallize membrane proteins." Protein Sci **18**(9): 1828-1839.

Kuiken, C., T. Leitner, B. Foley, B. Hahn, P. Marx, F. McCutchan, S. Wolinsky and B. Korber (2008). HIV Sequence Compendium 2008. T. L. Carla Kuiken, Brian Foley, Beatrice Hahn, Preston Marx, Francince McCutchan, Steven Wolinsky, Bette Korber. Los Alamos, Los Alamos National Laboratory.

Kwong, P. D., R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski and W. A. Hendrickson (1998). "Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody." Nature **393**(6686): 648-659.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J Mol Biol **157**(1): 105-132.

Land, A. M., T. B. Ball, M. Luo, R. Pilon, P. Sandstrom, J. E. Embree, C. Wachihi, J. Kimani and F. A. Plummer (2008). "Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya." J Virol **82**(16): 8172-8182.

Larder, B. A. and S. D. Kemp (1989). "Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT)." Science **246**(4934): 1155-1158.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

LaRue, R. S., V. Andresdottir, Y. Blanchard, S. G. Conticello, D. Derse, M. Emerman, W. C. Greene, S. R. Jonsson, N. R. Landau, M. Lochelt, H. S. Malik, M. H. Malim, C. Munk, S. J. O'Brien, V. K. Pathak, K. Strebel, S. Wain-Hobson, X. F. Yu, N. Yuhki and R. S. Harris (2009). "Guidelines for naming nonprimate APOBEC3 genes and proteins." J Virol **83**(2): 494-497.

Laskowski, R. A., J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-486.

Lecossier, D., F. Bouchonnet, F. Clavel and A. J. Hance (2003). "Hypermutation of HIV-1 DNA in the absence of the Vif protein." Science **300**(5622): 1112.

Lee, Y. N., M. H. Malim and P. D. Bieniasz (2008). "Hypermutation of an ancient human retrovirus by APOBEC3G." J Virol **82**(17): 8762-8770.

Lenassi, M., G. Cagney, M. Liao, T. Vaupotic, K. Bartholomeeusen, Y. Cheng, N. J. Krogan, A. Plemenitas and B. M. Peterlin (2010). "HIV Nef is secreted in exosomes and triggers apoptosis in bystander CD4+ T cells." Traffic **11**(1): 110-122.

Lesse, A. J., A. A. Campagnari, W. E. Bittner and M. A. Apicella (1990). "Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis." J Immunol Methods **126**(1): 109-117.

Levitt, M. H. (2008). Spin dynamics : basics of nuclear magnetic resonance. Chichester, England ; Hoboken, NJ, John Wiley & Sons.

Levy, D. N., Y. Refaeli, R. R. MacGregor and D. B. Weiner (1994). "Serum Vpr regulates productive infection and latency of human immunodeficiency virus type 1." Proc Natl Acad Sci U S A **91**(23): 10873-10877.

Liu, J., A. Bartesaghi, M. J. Borgnia, G. Sapiro and S. Subramaniam (2008). "Molecular architecture of native HIV-1 gp120 trimers." Nature **455**(7209): 109-113.

Lu, Z., W. Chen, R. Liu, X. Hu and Y. Ding (2010). "A novel method for high-level production of psychrophilic TAB5 alkaline phosphatase." Protein Expr Purif **74**(2): 217-222.

Luft, J. R., R. J. Collins, N. A. Fehrman, A. M. Lauricella, C. K. Veatch and G. T. DeTitta (2003). "A deliberate approach to screening for initial crystallization conditions of biological macromolecules." J Struct Biol **142**(1): 170-179.

Lundquist, C. A., M. Tobiume, J. Zhou, D. Unutmaz and C. Aiken (2002). "Nef-mediated downregulation

of CD4 enhances human immunodeficiency virus type 1 replication in primary T lymphocytes." J Virol **76**(9): 4625-4633.

Luo, K., Z. Xiao, E. Ehrlich, Y. Yu, B. Liu, S. Zheng and X. F. Yu (2005). "Primate lentiviral virion infectivity factors are substrate receptors that assemble with cullin 5-E3 ligase through a HCCH motif to suppress APOBEC3G." Proc Natl Acad Sci U S A **102**(32): 11444-11449.

Madani, N. and D. Kabat (1998). "An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein." J Virol **72**(12): 10251-10255.

Maertens, G. N., S. Hare and P. Cherepanov (2010). "The mechanism of retroviral integration from X-ray structures of its key intermediates." Nature **468**(7321): 326-329.

Mahrour, N., W. B. Redwine, L. Florens, S. K. Swanson, S. Martin-Brown, W. D. Bradford, K. Staehling-Hampton, M. P. Washburn, R. C. Conaway and J. W. Conaway (2008). "Characterization of Cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to Elongin BC-based ubiquitin ligases." J Biol Chem **283**(12): 8005-8013.

Malim, M. H. and M. Emerman (2008). "HIV-1 accessory proteins--ensuring viral survival in a hostile environment." Cell Host Microbe **3**(6): 388-398.

Mangeat, B., G. Gers-Huber, M. Lehmann, M. Zufferey, J. Luban and V. Piguet (2009). "HIV-1 Vpu neutralizes the antiviral factor Tetherin/BST-2 by binding it and directing its beta-TrCP2-dependent degradation." PLoS Pathog **5**(9): e1000574.

Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin and D. Trono (2003). "Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts." Nature **424**(6944): 99-103.

Marin, M., K. M. Rose, S. L. Kozak and D. Kabat (2003). "HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation." Nat Med **9**(11): 1398-1403.

Marion, D., P. C. Driscoll, L. E. Kay, P. T. Wingfield, A. Bax, A. M. Gronenborn and G. M. Clore (1989). "Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta." Biochemistry **28**(15): 6150-6156.

Markosyan, R. M., F. S. Cohen and G. B. Melikyan (2003). "HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation." Mol Biol Cell **14**(3): 926-938.

Mascola, J. R. and B. F. Haynes (2013). "HIV-1 neutralizing antibodies: understanding nature's

pathways." Immunol Rev **254**(1): 225-244.

McNatt, M. W., T. Zang and P. D. Bieniasz (2013). "Vpu binds directly to tetherin and displaces it from nascent virions." PLoS Pathog **9**(4): e1003299.

McNatt, M. W., T. Zang, T. Hatziioannou, M. Bartlett, I. B. Fofana, W. E. Johnson, S. J. Neil and P. D. Bieniasz (2009). "Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants." PLoS Pathog **5**(2): e1000300.

Mehle, A., J. Goncalves, M. Santa-Marta, M. McPike and D. Gabuzda (2004). "Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation." Genes Dev **18**(23): 2861-2866.

Mehle, A., B. Strack, P. Ancuta, C. Zhang, M. McPike and D. Gabuzda (2004). "Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway." J Biol Chem **279**(9): 7792-7798.

Mehle, A., E. R. Thomas, K. S. Rajendran and D. Gabuzda (2006). "A zinc-binding region in Vif binds Cul5 and determines cullin selection." J Biol Chem **281**(25): 17259-17265.

Meyer, B. E., J. L. Meinkoth and M. H. Malim (1996). "Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals." J Virol **70**(4): 2350-2359.

Michael, N. L. and J. P. Moore (1999). "HIV-1 entry inhibitors: evading the issue." Nat Med **5**(7): 740-742.

Michel, N., I. Allespach, S. Venzke, O. T. Fackler and O. T. Keppler (2005). "The Nef protein of human immunodeficiency virus establishes superinfection immunity by a dual strategy to downregulate cell-surface CCR5 and CD4." Curr Biol **15**(8): 714-723.

Mitsuya, H., K. J. Weinhold, P. A. Furman, M. H. St Clair, S. N. Lehrman, R. C. Gallo, D. Bolognesi, D. W. Barry and S. Broder (1985). "3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro." Proc Natl Acad Sci U S A **82**(20): 7096-7100.

Mulder, L. C., A. Harari and V. Simon (2008). "Cytidine deamination induced HIV-1 drug resistance." Proc Natl Acad Sci U S A **105**(14): 5501-5506.

National Institutes of Health (2002). International Committee on Taxonomy of Viruses. **61.0.6**.

Neil, S. J., T. Zang and P. D. Bieniasz (2008). "Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu." Nature **451**(7177): 425-430.

Nekhai, S. and K. T. Jeang (2006). "Transcriptional and post-transcriptional regulation of HIV-1 gene expression: role of cellular factors for Tat and Rev." Future Microbiol **1**(4): 417-426.

Nerdal, W., D. R. Hare and B. R. Reid (1989). "Solution structure of the EcoRI DNA sequence: refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations." Biochemistry **28**(26): 10008-10021.

Neuberger, M. S., R. S. Harris, J. Di Noia and S. K. Petersen-Mahrt (2003). "Immunity through DNA deamination." Trends Biochem Sci **28**(6): 305-312.

OhAinle, M., J. A. Kerns, M. M. Li, H. S. Malik and M. Emerman (2008). "Antiretroelement activity of APOBEC3H was lost twice in recent human evolution." Cell Host Microbe **4**(3): 249-259.

Oscar Millet, J. Patrick Loria, Christopher D. Kroenke, Miquel Pons and A. G. P. III (2000). "The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale." Journal of American Chemical Society **122**(12): 11.

Overhauser, A. W. (1953). "Polarization of Nuclei in Metals." Physical Review **92**(2): 5.

Pan, X., M. M. Geist, J. M. Rudolph, W. Nickel and O. T. Fackler (2013). "HIV-1 Nef disrupts membrane-microdomain-associated anterograde transport for plasma membrane delivery of selected Src family kinases." Cell Microbiol.

Pancera, M., S. Majeed, Y. E. Ban, L. Chen, C. C. Huang, L. Kong, Y. D. Kwon, J. Stuckey, T. Zhou, J. E. Robinson, W. R. Schief, J. Sodroski, R. Wyatt and P. D. Kwong (2010). "Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility." Proc Natl Acad Sci U S A **107**(3): 1166-1171.

Patick, A. K., H. Mo, M. Markowitz, K. Appelt, B. Wu, L. Musick, V. Kalish, S. Kaldor, S. Reich, D. Ho and S. Webber (1996). "Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease." Antimicrob Agents Chemother **40**(2): 292-297.

Paul, I., J. Cui and E. L. Maynard (2006). "Zinc binding to the HCCH motif of HIV-1 virion infectivity factor induces a conformational change that mediates protein-protein interactions." Proc Natl Acad Sci U S A **103**(49): 18475-18480.

Pauza, C. D., P. Trivedi, T. S. McKechnie, D. D. Richman and F. M. Graziano (1994). "2-LTR circular viral DNA as a marker for human immunodeficiency virus type 1 infection in vivo." Virology **205**(2): 470-478.

Pazhanisamy, S., C. M. Stuver, A. B. Cullinan, N. Margolin, B. G. Rao and D. J. Livingston (1996). "Kinetic characterization of human immunodeficiency virus type-1 protease-resistant variants." J Biol Chem

271(30): 17979-17985.

Peng, J., D. Schwartz, J. E. Elias, C. C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley and S. P. Gygi (2003). "A proteomics approach to understanding protein ubiquitination." Nat Biotechnol **21**(8): 921-926.

Petersen-Mahrt, S. K., R. S. Harris and M. S. Neuberger (2002). "AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification." Nature **418**(6893): 99-103.

Pham, P., R. Bransteitter and M. F. Goodman (2005). "Reward versus risk: DNA cytidine deaminases triggering immunity and disease." Biochemistry **44**(8): 2703-2715.

Pornillos, O., B. K. Ganser-Pornillos, B. N. Kelly, Y. Hua, F. G. Whitby, C. D. Stout, W. I. Sundquist, C. P. Hill and M. Yeager (2009). "X-ray structures of the hexameric building block of the HIV capsid." Cell **137**(7): 1282-1292.

Pornillos, O., B. K. Ganser-Pornillos and M. Yeager (2011). "Atomic-level modelling of the HIV capsid." Nature **469**(7330): 424-427.

Pornillos, O., J. E. Garrus and W. I. Sundquist (2002). "Mechanisms of enveloped RNA virus budding." Trends Cell Biol **12**(12): 569-579.

Qiao, X., B. He, A. Chiu, D. M. Knowles, A. Chadburn and A. Cerutti (2006). "Human immunodeficiency virus 1 Nef suppresses CD40-dependent immunoglobulin class switching in bystander B cells." Nat Immunol **7**(3): 302-310.

Rabi, I. I., S. Millman, P. Kusch and J. R. Zacharias (1939). "The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments." Physical Review **55**: 10.

Ramelot, T. A., S. Raman, A. P. Kuzin, R. Xiao, L. C. Ma, T. B. Acton, J. F. Hunt, G. T. Montelione, D. Baker and M. A. Kennedy (2009). "Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study." Proteins **75**(1): 147-167.

Reddy, T. R., H. Tang, W. Xu and F. Wong-Staal (2000). "Sam68, RNA helicase A and Tap cooperate in the post-transcriptional regulation of human immunodeficiency virus and type D retroviral mRNA." Oncogene **19**(32): 3570-3575.

Rice, L. M., T. N. Earnest and A. T. Brunger (2000). "Single-wavelength anomalous diffraction phasing revisited." Acta Crystallographica Section D-Biological Crystallography **56**: 1413-1420.

Richman, D. D., D. M. Margolis, M. Delaney, W. C. Greene, D. Hazuda and R. J. Pomerantz (2009). "The challenge of finding a cure for HIV infection." Science **323**(5919): 1304-1307.

Rizzuto, C. D., R. Wyatt, N. Hernandez-Ramos, Y. Sun, P. D. Kwong, W. A. Hendrickson and J. Sodroski (1998). "A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding." Science **280**(5371): 1949-1953.

Roeth, J. F. and K. L. Collins (2006). "Human immunodeficiency virus type 1 Nef: adapting to intracellular trafficking pathways." Microbiol Mol Biol Rev **70**(2): 548-563.

Rosenberg, B. R. and F. N. Papavasiliou (2007). "Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection." Adv Immunol **94**: 215-244.

Rotin, D. and S. Kumar (2009). "Physiological functions of the HECT family of ubiquitin ligases." Nat Rev Mol Cell Biol **10**(6): 398-409.

Russell, R. A. and V. K. Pathak (2007). "Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F." J Virol **81**(15): 8201-8210.

Saio, T., M. Yokochi, H. Kumeta and F. Inagaki (2010). "PCS-based structure determination of protein-protein complexes." J Biomol NMR **46**(4): 271-280.

Sauter, D. and F. Kirchhoff (2011). "Tetherin antagonism by primate lentiviral nef proteins." Curr HIV Res **9**(7): 514-523.

Sauter, D., D. Unterweger, M. Vogl, S. M. Usmani, A. Heigele, S. F. Kluge, E. Hermkes, M. Moll, E. Barker, M. Peeters, G. H. Learn, F. Bibollet-Ruche, J. V. Fritz, O. T. Fackler, B. H. Hahn and F. Kirchhoff (2012). "Human tetherin exerts strong selection pressure on the HIV-1 group N Vpu protein." PLoS Pathog **8**(12): e1003093.

Schagger, H. and G. von Jagow (1987). "Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa." Anal Biochem **166**(2): 368-379.

Schaller, T., C. Goujon and M. H. Malim (2012). "AIDS/HIV. HIV interplay with SAMHD1." Science **335**(6074): 1313-1314.

Schmitz, J. E., M. J. Kuroda, S. Santra, V. G. Sasseville, M. A. Simon, M. A. Lifton, P. Racz, K. Tenner-Racz, M. Dalesandro, B. J. Scallon, J. Ghrayeb, M. A. Forman, D. C. Montefiori, E. P. Rieber, N. L. Letvin and K. A. Reimann (1999). "Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes." Science **283**(5403): 857-860.

Schreiber, G. (2002). "Kinetic studies of protein-protein interactions." Curr Opin Struct Biol **12**(1): 41-47.

Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Schuler, W., K. Wecker, H. de Rocquigny, Y. Baudat, J. Sire and B. P. Roques (1999). "NMR structure of the (52-96) C-terminal domain of the HIV-1 regulatory protein Vpr: molecular insights into its biological functions." J Mol Biol **285**(5): 2105-2117.

Sheehy, A. M., N. C. Gaddis, J. D. Choi and M. H. Malim (2002). "Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein." Nature **418**(6898): 646-650.

Sheehy, A. M., N. C. Gaddis and M. H. Malim (2003). "The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif." Nat Med **9**(11): 1404-1407.

Shen, Y. and A. Bax (2007). "Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology." J Biomol NMR **38**(4): 289-302.

Shen, Y., F. Delaglio, G. Cornilescu and A. Bax (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR **44**(4): 213-223.

Shen, Y., O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker and A. Bax (2008). "Consistent blind protein structure generation from NMR chemical shift data." Proc Natl Acad Sci U S A **105**(12): 4685-4690.

Shen, Y., R. Vernon, D. Baker and A. Bax (2009). "De novo protein structure generation from incomplete chemical shift assignments." J Biomol NMR **43**(2): 63-78.

Siliciano, J. D., J. Kajdas, D. Finzi, T. C. Quinn, K. Chadwick, J. B. Margolick, C. Kovacs, S. J. Gange and R. F. Siliciano (2003). "Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells." Nat Med **9**(6): 727-728.

Simon, J. H., N. C. Gaddis, R. A. Fouchier and M. H. Malim (1998). "Evidence for a newly discovered cellular anti-HIV-1 phenotype." Nat Med **4**(12): 1397-1400.

Simon, J. H., A. M. Sheehy, E. A. Carpenter, R. A. Fouchier and M. H. Malim (1999). "Mutational analysis of the human immunodeficiency virus type 1 Vif protein." J Virol **73**(4): 2675-2681.

Sloan, R. D., D. A. Donahue, B. D. Kuhl, T. Bar-Magen and M. A. Wainberg (2010). "Expression of Nef from unintegrated HIV-1 DNA downregulates cell surface CXCR4 and CCR5 on T-lymphocytes." Retrovirology **7**: 44.

Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger and W. Haseltine (1986). "A second post-transcriptional trans-activator gene required for HTLV-III replication." Nature **321**(6068): 412-417.

Stanley, B. J., E. S. Ehrlich, L. Short, Y. Yu, Z. Xiao, X. F. Yu and Y. Xiong (2008). "Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly." J Virol **82**(17): 8656-8663.

Stephenson, K. E. and D. H. Barouch (2013). "A global approach to HIV-1 vaccine development." Immunol Rev **254**(1): 295-304.

Su, X. C. and G. Otting (2010). "Paramagnetic labelling of proteins and oligonucleotides for NMR." J Biomol NMR **46**(1): 101-112.

Swanstrom, R. and J. W. Wills (1997). "Synthesis, Assembly, and Processing of Viral Proteins."

Tahirov, T. H., N. D. Babayeva, K. Varzavand, J. J. Cooper, S. C. Sedore and D. H. Price (2010). "Crystal structure of HIV-1 Tat complexed with human P-TEFb." Nature **465**(7299): 747-751.

Tan, L., P. T. Sarkis, T. Wang, C. Tian and X. F. Yu (2009). "Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1." FASEB J **23**(1): 279-287.

Techtmann, S. M., R. Ghirlando, S. Kao, K. Strebel and E. L. Maynard (2012). "Hydrodynamic and functional analysis of HIV-1 Vif oligomerization." Biochemistry **51**(10): 2078-2086.

Tjandra, N., D. S. Garrett, A. M. Gronenborn, A. Bax and G. M. Clore (1997). "Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy." Nat Struct Biol **4**(6): 443-449.

Tjandra, N., J. G. Omichinski, A. M. Gronenborn, G. M. Clore and A. Bax (1997). "Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution." Nat Struct Biol **4**(9): 732-738.

Tolman, J. R., J. M. Flanagan, M. A. Kennedy and J. H. Prestegard (1995). "Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution." Proc Natl Acad Sci U S A **92**(20): 9279-9283.

Tomasicchio, M., C. Avenant, A. Du Toit, R. M. Ray and J. P. Hapgood (2013). "The Progestin-Only Contraceptive Medroxyprogesterone Acetate, but Not Norethisterone Acetate, Enhances HIV-1 Vpr-Mediated Apoptosis in Human CD4(+) T Cells through the Glucocorticoid Receptor." PLoS One **8**(5): e62895.

Tristem, M., A. Purvis and D. L. Quicke (1998). "Complex evolutionary history of primate lentiviral vpr genes." Virology **240**(2): 232-237.

Trono, D., C. Van Lint, C. Rouzioux, E. Verdin, F. Barre-Sinoussi, T. W. Chun and N. Chomont (2010). "HIV

persistence and the prospect of long-term drug-free remissions for HIV-infected individuals." Science **329**(5988): 174-180.

Tu, X., K. Das, Q. Han, J. D. Bauman, A. D. Clark, Jr., X. Hou, Y. V. Frenkel, B. L. Gaffney, R. A. Jones, P. L. Boyer, S. H. Hughes, S. G. Sarafianos and E. Arnold (2010). "Structural basis of HIV-1 resistance to AZT by excision." Nat Struct Mol Biol **17**(10): 1202-1209.

Tzeng, S. R. and C. G. Kalodimos (2012). "Protein activity regulation by conformational entropy." Nature **488**(7410): 236-240.

Van Damme, N., D. Goff, C. Katsura, R. L. Jorgenson, R. Mitchell, M. C. Johnson, E. B. Stephens and J. Guatelli (2008). "The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein." Cell Host Microbe **3**(4): 245-252.

Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue (2005). "The CCPN data model for NMR spectroscopy: development of a software pipeline." Proteins **59**(4): 687-696.

Walker, J. M. (2005). The proteomics protocols handbook. Totowa, N.J., Humana Press.

Walker, R. C., Jr., M. A. Khan, S. Kao, R. Goila-Gaur, E. Miyagi and K. Strebel (2010). "Identification of dominant negative human immunodeficiency virus type 1 Vif mutants that interfere with the functional inactivation of APOBEC3G by virus-encoded Vif." J Virol **84**(10): 5201-5211.

Wang, X., S. Singh, H. Y. Jung, G. Yang, S. Jun, K. J. Sastry and J. I. Park (2013). "HIV-1 Vpr Inhibits Telomerase Activity Via EDD-DDB1-VPRBP E3 Ligase Complex." J Biol Chem.

Wassenaar, T. A., M. van Dijk, N. Loureiro-Ferreira, G. van der Schot, S. J. de Vries, C. Schmitz, J. van der Zwan, R. Boelens, A. Giachetti, L. Ferella, A. Rosato, I. Bertini, T. Herrmann, H. R. A. Jonker, A. Bagaria, V. Jaravine, P. Guntert, H. Schwalbe, W. F. Vranken, J. F. Doreleijers, G. Vriend, G. W. Vuister, D. Franke, A. Kikhney, D. I. Svergun, R. H. Fogh, J. Ionides, E. D. Laue, C. Spronk, S. Jurksa, M. Verlato, S. Badoer, S. Dal Pra, M. Mazzucato, E. Frizziero and A. M. J. J. Bonvin (2012). "WeNMR: Structural Biology on the Grid." Journal of Grid Computing **10**(4): 743-767.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature **171**(4356): 737-738.

Wei, W., H. Guo, X. Han, X. Liu, X. Zhou, W. Zhang and X. F. Yu (2012). "A novel DCAF1-binding motif required for Vpx-mediated degradation of nuclear SAMHD1 and Vpr-induced G2 arrest." Cell Microbiol **14**(11): 1745-1756.

WHO, U. (2010). "Global Report."

Wiegand, H. L., B. P. Doehle, H. P. Bogerd and B. R. Cullen (2004). "A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins." EMBO J **23**(12): 2451-2458.

Wilk, T., I. Gross, B. E. Gowen, T. Rutten, F. de Haas, R. Welker, H. G. Krausslich, P. Boulanger and S. D. Fuller (2001). "Organization of immature human immunodeficiency virus type 1." J Virol **75**(2): 759-771.

Wissing, S., N. L. Galloway and W. C. Greene (2010). "HIV-1 Vif versus the APOBEC3 cytidine deaminases: an intracellular duel between pathogen and host restriction factors." Mol Aspects Med **31**(5): 383-397.

Wittlich, M., B. W. Koenig, M. Stoldt, H. Schmidt and D. Willbold (2009). "NMR structural characterization of HIV-1 virus protein U cytoplasmic domain in the presence of dodecylphosphatidylcholine micelles." FEBS J **276**(22): 6560-6575.

Wolfe, L. S., B. J. Stanley, C. Liu, W. K. Eliason and Y. Xiong (2010). "Dissection of the HIV Vif interaction with human E3 ubiquitin ligase." J Virol **84**(14): 7135-7139.

Wood, N., T. Bhattacharya, B. F. Keele, E. Giorgi, M. Liu, B. Gaschen, M. Daniels, G. Ferrari, B. F. Haynes, A. McMichael, G. M. Shaw, B. H. Hahn, B. Korber and C. Seoighe (2009). "HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC." PLoS Pathog **5**(5): e1000414.

Wu, X., Z. Y. Yang, Y. Li, C. M. Hogerkorp, W. R. Schief, M. S. Seaman, T. Zhou, S. D. Schmidt, L. Wu, L. Xu, N. S. Longo, K. McKee, S. O'Dell, M. K. Louder, D. L. Wycuff, Y. Feng, M. Nason, N. Doria-Rose, M. Connors, P. D. Kwong, M. Roederer, R. T. Wyatt, G. J. Nabel and J. R. Mascola (2010). "Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1." Science **329**(5993): 856-861.

Wu, X., T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N. S. Longo, M. Louder, K. McKee, S. O'Dell, S. Perfetto, S. D. Schmidt, W. Shi, L. Wu, Y. Yang, Z. Y. Yang, Z. Yang, Z. Zhang, M. Bonsignori, J. A. Crump, S. H. Kapiga, N. E. Sam, B. F. Haynes, M. Simek, D. R. Burton, W. C. Koff, N. A. Doria-Rose, M. Connors, J. C. Mullikin, G. J. Nabel, M. Roederer, L. Shapiro, P. D. Kwong and J. R. Mascola (2011). "Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing." Science **333**(6049): 1593-1602.

Xing, S., S. Bhat, N. S. Shroff, H. Zhang, J. A. Lopez, J. B. Margolick, J. O. Liu and R. F. Siliciano (2012). "Novel structurally related compounds reactivate latent HIV-1 in a bcl-2-transduced primary CD4+ T cell model without inducing global T cell activation." J Antimicrob Chemother **67**(2): 398-403.

Xu, H., E. Chertova, J. Chen, D. E. Ott, J. D. Roser, W. S. Hu and V. K. Pathak (2007). "Stoichiometry of the

antiviral protein APOBEC3G in HIV-1 virions." Virology **360**(2): 247-256.

Xu, P. and J. Peng (2008). "Characterization of polyubiquitin chain structure by middle-down mass spectrometry." Anal Chem **80**(9): 3438-3444.

Yan, Q., T. Kamura, Y. Cai, J. Jin, M. Ivan, A. Mushegian, R. C. Conaway and J. W. Conaway (2004). "Identification of Elongin C and Skp1 sequences that determine Cullin selection." J Biol Chem **279**(41): 43019-43026.

Yang, B., L. Gao, L. Li, Z. Lu, X. Fan, C. A. Patel, R. J. Pomerantz, G. C. DuBois and H. Zhang (2003). "Potent suppression of viral infectivity by the peptides that inhibit multimerization of human immunodeficiency virus type 1 (HIV-1) Vif proteins." J Biol Chem **278**(8): 6596-6602.

Yang, H. C., S. Xing, L. Shan, K. O'Connell, J. Dinoso, A. Shen, Y. Zhou, C. K. Shrum, Y. Han, J. O. Liu, H. Zhang, J. B. Margolick and R. F. Siliciano (2009). "Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation." J Clin Invest **119**(11): 3473-3486.

Yang, S., Y. Sun and H. Zhang (2001). "The multimerization of human immunodeficiency virus type I Vif protein: a requirement for Vif function in the viral life cycle." J Biol Chem **276**(7): 4889-4893.

Yeager, M., E. M. Wilson-Kubalek, S. G. Weiner, P. O. Brown and A. Rein (1998). "Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms." Proc Natl Acad Sci U S A **95**(13): 7299-7304.

Yedavalli, V. S., C. Neuveut, Y. H. Chi, L. Kleiman and K. T. Jeang (2004). "Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function." Cell **119**(3): 381-392.

Yu, X., Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao and X. F. Yu (2003). "Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex." Science **302**(5647): 1056-1060.

Yu, Y., Z. Xiao, E. S. Ehrlich, X. Yu and X. F. Yu (2004). "Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines." Genes Dev **18**(23): 2867-2872.

Yu, Z., N. Sanchez-Velar, I. E. Catrina, E. L. Kittler, E. B. Udofia and M. L. Zapp (2005). "The cellular HIV-1 Rev cofactor hRIP is required for viral replication." Proc Natl Acad Sci U S A **102**(11): 4027-4032.

Zadjali, F., A. C. Pike, M. Vesterlund, J. Sun, C. Wu, S. S. Li, L. Ronnstrand, S. Knapp, A. N. Bullock and A. Flores-Morales (2011). "Structural basis for c-KIT inhibition by the suppressor of cytokine signaling 6 (SOCS6) ubiquitin ligase." J Biol Chem **286**(1): 480-490.

Zhang, F., S. J. Wilson, W. C. Landford, B. Virgen, D. Gregory, M. C. Johnson, J. Munch, F. Kirchhoff, P. D. Bieniasz and T. Hatziioannou (2009). "Nef proteins from simian immunodeficiency viruses are tetherin antagonists." Cell Host Microbe **6**(1): 54-67.

Zhang, J. G., A. Farley, S. E. Nicholson, T. A. Willson, L. M. Zugaro, R. J. Simpson, R. L. Moritz, D. Cary, R. Richardson, G. Hausmann, B. J. Kile, S. B. Kent, W. S. Alexander, D. Metcalf, D. J. Hilton, N. A. Nicola and M. Baca (1999). "The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation." Proc Natl Acad Sci U S A **96**(5): 2071-2076.

Zhang, W., J. Du, S. L. Evans, Y. Yu and X. F. Yu (2012). "T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction." Nature **481**(7381): 376-379.

Zhao, G., J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken and P. Zhang (2013). "Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics." Nature **497**(7451): 643-646.

Zhao, K., Y. Ishida, T. K. Oleksyk, C. A. Winkler and A. L. Roca (2012). "Evidence for selection at HIV host susceptibility genes in a West Central African human population." BMC Evol Biol **12**: 237.

Zheng, N., B. A. Schulman, L. Song, J. J. Miller, P. D. Jeffrey, P. Wang, C. Chu, D. M. Koepp, S. J. Elledge, M. Pagano, R. C. Conaway, J. W. Conaway, J. W. Harper and N. P. Pavletich (2002). "Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex." Nature **416**(6882): 703-709.

Zheng, Y. H., D. Irwin, T. Kurosu, K. Tokunaga, T. Sata and B. M. Peterlin (2004). "Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication." J Virol **78**(11): 6073-6076.

Zhou, X., S. L. Evans, X. Han, Y. Liu and X. F. Yu (2012). "Characterization of the interaction of full-length HIV-1 Vif protein with its key regulator CBFbeta and CRL5 E3 ubiquitin ligase components." PLoS One **7**(3): e33495.

Zhu, P., J. Liu, J. Bess, Jr., E. Chertova, J. D. Lifson, H. Grise, G. A. Ofek, K. A. Taylor and K. H. Roux (2006). "Distribution and three-dimensional structure of AIDS virus envelope spikes." Nature **441**(7095): 847-852.

APPENDICES:

1. Primer used for PCR

Name	Sequence	Ter	Gene	Comments
N_SET_VIF	GGCCATGGGCCAGTACAAACTGATCCTG	5'	Vif	For SET-Vif cloning
C_SET_VIF	GGGTCGACGAATTCTTAGTGGTGGTGGTG	3'	Vif	For SET-Vif cloning
SET_Vif_Cstop	GGAAGCTTGTCGACTTAGTCGACCTAGTG	3'	Vif	For SET-Vif cloning
SET_Vif_NHis_SET	GGCTCTAGCCATGGGCCACCATCACCATCACCACCATATGCAGTACAAACTG	5'	His-SET	For His-SET-Vif cloning
N_EcoRI-SacI-SET	GGGAATTCGAGCTCGATGCAGTACAAAC	5'	SET	
C_CBFb1t_XhoI	GGctcgagCTAGGGTCTTGTTGTCTTCTT	3'	CBFb-II	For CBFb isoform I (shorter Isoform)
N_His_CBFb	GGcatatgCATCACCATCACCACACATGCCGCGCGTCGTGCCCGACCAG	5'	CBFb	For His-CBFb cloning
C_CBFb2o_XhoI	GGctcgagTTAACGCAGTTTCAGGTC	3'	CBFb-I	For CBFb isoform II (longer Isoform)
SET_Vif_Mut_N	GCCATACAATGAATGAATGGACACTAG	5'	Vif	A mutation on original tempalte
SET_Vif_Mut_C	CTAGTGTCCATTCATTCATTGTATGGC	3'	Vif	A mutation on original tempalte
NcoI-EloB	GGCCATGGACGTGTTCCTCATGATCCG	5'	EloB	
DuetDOWN1_Primer	GGGATTATGCGGCCGTGTACAA	3'	-	Duet Vector 3' primer
N_His_CBFb	GGcatatgCATCACCATCACCACCACATGCCGCGCGTCGTG	5'	CBFb	High pure
XhoI_T7_promoter	GGACCCTCGAGGAAATTAATACGACTCAC	5'	-	general T7 primer
T7-term primer	CCGCTAGTTATTGCTCAGCGGT	3'	-	general T7 terminal primer
N_Pep5_L165A	GCCACCTGCCCCTAGTGTTAGGAAACTG	5'	Vif	L163A mutation, failed
C_Pep5_L165A	CCTAACACTAGGGGCAGGTGGCTTTATC	3'	Vif	L163A mutation, failed
N_Pep5_S168A	GCCACCTTTGCCTGCTGTTAGGAAACTG	5'	Vif	S165A mutation for ITC studies

C_Pep5_S168A	CCTAACAGCAGGCAAAGGTGGCTTTATC	3'	Vif	S165A mutation for ITC studies
SPLP	AAAACCAAAACAGATAAAGAGCCCTTTGCCTAGTGTTAGGA	5'	Vif	P161S mutation for ITC studies
SPLP	TCCTAACACTAGGCAAAGGGCTCTTTATCTGTTTTGGTTTT	3'	Vif	P161S mutation for ITC studies
PSLP	ACCAAAACAGATAAAGCCAAGCTTGCCTAGTGTTAGGAAAC	5'	Vif	P162S mutation for ITC studies
PSLP	GTTTCCTAACACTAGGCAAGCTTGGCTTTATCTGTTTTGGT	3'	Vif	P162S mutation for ITC studies
PPLS	ACAGATAAAGCCACCTTTGAGCAGTGTTAGGAAACTGACAG	5'	Vif	P164S mutation for ITC studies
PPLS	CTGTCAGTTTCCTAACACTGCTCAAAGGTGGCTTTATCTGT	3'	Vif	P164S mutation for ITC studies
SET_Vif_Mut_N	GCCATACAATGAATGGACACTAGGTCG	5'	Vif	
SET_Vif_Mut_C	CCTAGTGTCCATTCATTGTATGGCTCC	3'	Vif	
Vif_C-term_stop	CCCAAGCTTGTCGACTCAGTGTCCATTC	3'	Vif	
NcoI_HIV2Vif	GGGAAACCATGGAGGAAGACAAGAGATG	5'	HIV-2 Vif	
HIV2Vif_stop_C	CCGCAAGCTTGTCGACTCATGCCAGTATC	3'	HIV-2 Vif	
N_BamHI_SOCS	GGGGAAAGGATCCCAGGTACCGTCACTTC	5'	Vif	
Vif_His_end_C	CCAAGCTTGTCGACGAATTCTTAGTGGTGGTGGTGGTGTGCCAGTATCTCCAGGAC	3'	Vif	For Vif-His
SOCS_His_end_C	CCAAGCTTGTCGACGAATTCTTAGTGGTGGTGGTGGTGTCTCCGATAGTCTCTTCGCC	3'	Vif	For SOCS-His
N_Pep5_G143C	CATAACAAGGTATGTTCTCTACAGTAC	5'	Vif	PRE studies
C_Pep5_G143C	TGTAGAGAACATACCTTGTTATGGGATCC	3'	Vif	PRE studies
N_Pep5_Q158C	CCAAAATGTATAAAGCCACCTTTG	5'	Vif	PRE studies
C_Pep5_Q158C	GGCTTTATACATTTTGGTTTTATTAATG	3'	Vif	PRE studies
N_Pep5_R167C	CCTAGTGTTTGTAAACTGACAGAGGAC	5'	Vif	PRE studies
C_Pep5_R167C	CTGTCAGTTTACAAACACTAGGCAAAGG	3'	Vif	PRE studies
N_Vif_L163F	GATAAAGCCACCTTTTCCTAGTGTTAGGAAACTG	5'	Vif	For biological studies
C_Vif_L163F	CAGTTTCCTAACACTAGGAAAAGGTGGCTTTATC	3'	Vif	For biological studies
N_EloB_D101A	CCCGCCAGAGCTGCCCGCTGTGATGAAGCCCCAGG	5'	EloB	NMR Perturbation studies
C_EloB_D101A	CTGGGGCTTCATCACAGCGGGCAGCTCTGGCGGGC	3'	EloB	NMR Perturbation studies

N_EloB_V102A	CCAGAGCTGCCCGATGCGATGAAGCCCCAGGACTC	5'	EloB	NMR Perturbation studies
C_EloB_V102A	GTCCTGGGGCTTCATCGCATCGGGCAGCTCTGGC	3'	EloB	NMR Perturbation studies
N_EloB_M103A	GAGCTGCCCGATGTGGCGAAGCCCCAGGACTCGGG	5'	EloB	NMR Perturbation studies
C_EloB_M103A	CCCGAGTCCTGGGGCTTCGCCACATCGGGCAGCTC	3'	EloB	NMR Perturbation studies
N_EloB_K104A	CTGCCCGATGTGATGGCGCCCCAGGACTCGGGAAG	5'	EloB	NMR Perturbation studies
C_EloB_K104A	CTTCCCGAGTCCTGGGGCGCCATCACATCGGGCAG	3'	EloB	NMR Perturbation studies

2. Scripts for amino acid statistics

```
#! /usr/bin/perl -w

"<$ARGV[0]";
print "Reading file: $ARGV[0] ... \n";
sleep 1;
open (FILE, "$ARGV[0]") or die "Cannot read the file!\n";

@line=<FILE>;
chomp(@line);
$l=0;
$n=0;

my $A=0;
my $C=0;
my $D=0;
my $E=0;
my $F=0;
my $G=0;
my $H=0;
my $I=0;
my $K=0;
my $L=0;
my $M=0;
my $N=0;
my $P=0;
my $Q=0;
my $R=0;
my $S=0;
my $T=0;
my $V=0;
my $W=0;
my $Y=0;

foreach $line(@line){
     if ($l == 0){
          $name=substr($line, 1, 100);
#         $name_length=length($name);
#         print "$name, $name_length\n";
          $l++;
          }
     else{
```

```
        chomp($line);
        $wholeseq=$wholeseq . "$line";
        @aa=split(//,$line);
        foreach $aa(@aa){
            use 5.010;
            given ($aa){
                when ($aa eq A) {$A++;$n++;push (@water,"~");$weight=$weight+ 71.05}
                when ($aa eq C) {$C++;$n++;push (@water,"n");$weight=$weight+103.07}
                when ($aa eq D) {$D++;$n++;push (@water,"p");$weight=$weight+115.06}
                when ($aa eq E) {$E++;$n++;push (@water,"p");$weight=$weight+129.08}
                when ($aa eq F) {$F++;$n++;push (@water,"~");$weight=$weight+147.12}
                when ($aa eq G) {$G++;$n++;push (@water,"n");$weight=$weight+ 57.03}
                when ($aa eq H) {$H++;$n++;push (@water,"p");$weight=$weight+137.10}
                when ($aa eq I) {$I++;$n++;push (@water,"~");$weight=$weight+113.12}
                when ($aa eq K) {$K++;$n++;push (@water,"p");$weight=$weight+128.13}
                when ($aa eq L) {$L++;$n++;push (@water,"~");$weight=$weight+113.12}
                when ($aa eq M) {$M++;$n++;push (@water,"~");$weight=$weight+131.12}
                when ($aa eq N) {$N++;$n++;push (@water,"n");$weight=$weight+114.07}
                when ($aa eq P) {$P++;$n++;push (@water,"~");$weight=$weight+ 97.08}
                when ($aa eq Q) {$Q++;$n++;push (@water,"n");$weight=$weight+128.09}
                when ($aa eq R) {$R++;$n++;push (@water,"p");$weight=$weight+156.14}
                when ($aa eq S) {$S++;$n++;push (@water,"n");$weight=$weight+ 87.05}
                when ($aa eq T) {$T++;$n++;push (@water,"n");$weight=$weight+101.07}
                when ($aa eq V) {$V++;$n++;push (@water,"~");$weight=$weight+ 99.10}
                when ($aa eq W) {$W++;$n++;push (@water,"~");$weight=$weight+186.15}
                when ($aa eq Y) {$Y++;$n++;push (@water,"n");$weight=$weight+163.12}
                }
            }
            $l++;
        }
    }

$aminoacid{"A"}=$A;
$aminoacid{"C"}=$C;
$aminoacid{"D"}=$D;
$aminoacid{"E"}=$E;
$aminoacid{"F"}=$F;
$aminoacid{"G"}=$G;
$aminoacid{"H"}=$H;
$aminoacid{"I"}=$I;
$aminoacid{"K"}=$K;
$aminoacid{"L"}=$L;
$aminoacid{"M"}=$M;
$aminoacid{"P"}=$P;
```

```
$aminoacid{"Q"}=$Q;
$aminoacid{"R"}=$R;
$aminoacid{"S"}=$S;
$aminoacid{"T"}=$T;
$aminoacid{"V"}=$V;
$aminoacid{"W"}=$W;
$aminoacid{"Y"}=$Y;

$all=($A+$C+$D+$E+$F+$G+$H+$I+$K+$L+$M+$N+$P+$Q+$R+$S+$T+$V+$W+$Y);
print "\n>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>\n";
if ($n == $all){
    print " Protein $name has $all residues. The molecular weight is $weight Da.\n";
    my @allres = sort by_number keys %aminoacid;
    sub by_number {$aminoacid{$b} <=> $aminoacid{$a}}
    print " \n The descending order of frequency of each amino acid is: \n  @allres
\n\n The details are: \n";
    while (($res, $value) = each %aminoacid){
        printf "  %-2s residues in protein $name are %1s.\n", $value, $res;
    }
    print "\n The hydrophobic / hydrophilic characters of the sequence is: \n";
    foreach $water(@water){
        print "$water";
    }
    print "\n *'p' stands for polor hydrophilic.\n  'n' stands for non-polor
hydrophilic.\n  '~' stands for hydrophobic.\n ";
    $_=$wholeseq;
    use 5.010;
    given($_){
        when(/LVPRGS/i)  {print "\n Warning! There is a thrombin cleavage site in this
sequence!\n"; continue}
        when(/ENLYFQG/i) {print "\n Warning! There is a TEV cleavage site in this
sequence!\n"; continue}
        when(/LEVLFQGP/i){print "\n Warning! There is a PreScission cleavage site in
this sequence!\n"; continue}
    }
}
else {
    print "\nWrong! Please check the file!\n";
}
print "\n-> p stands for polar residue, n stands for non-polar hydrophilic residue and
~ stands for hydorphobic residue.\n";
print "\n--END--\n";
```

3. The assignments of Vif SOCS-box and EloBC

The assignments of Vif SOCS-box and EloBC are completed by Dr. Julien Bergeron and me, and deposited to BMRB database, ID: 19333. The full list of assignments is attached in the electronic version of this thesis.

4. Vif amino acid sequence alignment

Vif sequences selected from Los Alamos HIV-1 sequence database are aligned and attached in the electronic version of this thesis.

5. The files of mass spectrometry

The detail of the experiment is described in Chapter 2. Two raw data files and a reported offered by KCL Proteomics Facility, Institute of Psychiatry are attached in the electronic version of this thesis.

Mark	Comments
c	clear
?	odd, contamination perhaps
mp	minus precipitation
p	precipitation
salt	salt crystals
s/p	salt crystals and precipitation
X	drops were demaged

Date 20121218
Protein Vif-EloBC-CBFb
[con] 4 mgs/ml **Method** BioRed

Salt 250 mM Nacl
Buffer 50 mM Tris

Drop1 condition 1 ul sample 1 ul
Drop2 condition 1 ul sample 2 ul

Prep. 50% Ammonium sulphate 50% Methonal
Cell: pH gradient

	1	2	3	4	5	6
A	3	5.5	7.5	4	6	8
B	3.5	6	8	4.5	6.5	8.5
C	4	6.5	8.5	5	7	9
D	4.5	7	9	5.5	7.5	9.5

Check 1 20121219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130104

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130117

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 5 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Date 20121218
Protein Vif-EloBC-CBFb
[con] 4 mgs/ml **Method** BioRed

Salt 250 mM Nacl
Buffer 50 mM Tris

Drop1 condition 1 ul sample 1 ul
Drop2 condition 1 ul sample 2 ul

Prep. 40% PEG 400 10% PEG 6000
Cell: pH gradient

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20121219

	1	2	3	4	5	6
A	p	p	c	p	p	p
B	p	p	c	p	p	p
C	p	c	c	p	p	p
D	p	c	c	p	p	p

Check 2 20130104

	1	2	3	4	5	6
A	p	p	mp	p	p	p
B	p	p	mp	p	p	p
C	p	mp	c	p	p	p
D	p	mp	c	p	p	p

Check 3 20130117

	1	2	3	4	5	6
A	p	p	mp	p	p	p
B	p	p	mp	p	p	p
C	p	p	c	p	p	p
D	p	mp	c	p	p	p

Check 4 20130202

	1	2	3	4	5	6
A	p	p	mp	p	p	p
B	p	p	mp	p	p	p
C	p	p	?	p	p	p
D	p	p	c	p	p	p

Check 5 20130219

	1	2	3	4	5	6
A	p	p	mp	p	p	p
B	p	p	mp	p	p	p
C	p	p	?	p	p	p
D	p	p	c	p	p	p

Check 5 20130307

	1	2	3	4	5	6
A	p	p	mp	p	p	p

B	p	p	mp	p	p	p
C	p	p	?	p	p	p
D	p	p	c	p	p	p

Check 6 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	c	p	p	p
D	p	p	c	p	p	p

Check 7 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	X	p	p	p
D	p	p	X	p	p	p

Date 20121219
Protein Vif-EloBC-CBFb
[con] 3.5 mg/ml **Method** BioRed

Salt 250 mM Nacl
Buffer 50 mM Tris

Drop1 condition 1 ul sample 1 ul
Drop2 condition 1 ul sample 2 ul

Prep. 25% Ammonium sulphate 10% Ammonium sulphate
Cell: pH gradient

	1	2	3	4	5	6
A	3	5.5	7.5	3	5.5	7.5
B	3.5	6	8	3.5	6	8
C	4	6.5	8.5	4	6.5	8.5
D	4.5	7	9	4.5	7	9

Check 1 20130104

	1	2	3	4	5	6
A	mp	c	c	mp	c	c
B	mp	mp	c	c	c	c
C	mp	c	c	c	c	c
D	mp	c	c	c	c	c

Check 2 20130117

	1	2	3	4	5	6
A	p	mp	p	mp	c	c
B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	mp	p	p	c	c	c

Check 3 20130202

	1	2	3	4	5	6
A	p	p	p	mp	c	c
B	p	p	p	mp	c	?
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 4 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	c
B	p	p	p	p	c	?
C	p	p	p	mp	c	c
D	p	p	p	mp	c	mp

Check 5 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	c
B	p	p	p	p	c	?
C	p	p	p	mp	c	c
D	p	p	p	mp	c	mp

Check 6 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	p

B	p	p	p	p	p	p
C	p	p	p	mp	mp	p
D	p	p	p	mp	mp	p

Date 20121219
Protein Vif-EloBC-CBFb
[con] 3.5 mg/ml **Method** BioRed

Salt 250 mM Nacl
Buffer 50 mM Tris

Drop1 condition 1 ul sample 1 ul
Drop2 condition 1 ul sample 2 ul

Prep. 25% Methonal 10% Methonal
Cell: pH gradient

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130104

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	c	c
C	mp	mp	p	c	c	c
D	p	p	p	c	c	c

Check 2 20130117

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 3 20130202

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 4 20130219

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 5 20130307

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 6 20130314

	1	2	3	4	5	6
A	p	p	p	c	c	c

B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 6 20130409

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	c	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Date 20121219
Protein Vif-EloBC-CBFb
[con] 3.5 mg/ml **Method** BioRed

Salt 250 mM Nacl
Buffer 50 mM Tris

Drop1 condition 1 ul sample 1 ul
Drop2 condition 1 ul sample 2 ul

Prep. 20%, *60%* PEG 400 10%, *80%* PEG 400
Cell: pH gradient

	1	2	3	4	5	6
A	4	6	*8*	4	6	*8*
B	4.5	6.5	*8.5*	4.5	6.5	*8.5*
C	5	*7*	*9*	5	*7*	*9*
D	5.5	*7.5*	*9.5*	5.5	*7.5*	*9.5*

Check 1 20130104

	1	2	3	4	5	6
A	mp	c	p	c	c	c
B	c	c	p	c	c	c
C	c	p	p	c	mp	c
D	c	p	p	c	c	c

Check 2 20130117

	1	2	3	4	5	6
A	mp	mp	p	p	p	mp
B	mp	c	p	p	p	c
C	mp	p	p	p	p	c
D	c	p	p	p	mp	c

Check 3

	1	2	3	4	5	6
A	p	p	p	p	p	mp
B	p	c	p	p	p	c
C	p	p	p	p	p	c
D	mp	p	p	p	p	c

Check 4 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	mp
B	p	c	p	p	p	mp
C	p	p	p	p	p	mp
D	p	p	p	p	p	mp

Check 5 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	mp
B	p	c	p	p	p	mp
C	p	p	p	p	p	mp
D	p	p	p	p	p	mp

Check 5 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	mp

B	p	c	p	p	p	mp
C	p	p	p	p	p	mp
D	p	p	p	p	p	mp

Check 6 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	mp
B	p	p	p	p	p	mp
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Date 20121219
Protein Vif-EloBC-CBFb
[con] 3.5 mg/ml **Method** BioRed

Salt 250 mM Nacl
Buffer 50 mM Tris

Drop1 condition 1 ul sample 1 ul
Drop2 condition 1 ul sample 2 ul

Prep. 5% PEG 6000 2.5% PEG 6000
Cell: pH gradient

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130104

	1	2	3	4	5	6
A	p	p	p	c	c	c
B	p	p	p	c	X	c
C	p	p	p	c	c	c
D	p	p	p	c	c	c

Check 2 20130117

	1	2	3	4	5	6
A	p	p	p	c	mp	mp
B	p	p	p	c	X	c
C	p	p	p	c	mp	mp
D	p	p	p	c	c	c

Check 3

	1	2	3	4	5	6
A	p	p	p	mp	mp	mp
B	p	p	p	mp	X	mp
C	p	p	p	mp	mp	mp
D	p	p	p	mp	mp	mp

Check 4 20130219

	1	2	3	4	5	6
A	p	p	p	mp	mp	p
B	p	p	p	mp	X	X
C	p	p	p	mp	p	p
D	p	p	p	mp	p	p

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 0.2 M CaCl
Buffer 0.1 M HEPES
Prep. 40% PEG 400

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.		no Zn			1 mM Zn	
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130125

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 0.1 M MgCl2
Buffer 0.1 M Tris
Prep. 8% PEG 8000

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.		no Zn			1 mM Zn	
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130125

	1	2	3	4	5	6
A	mp	mp	mp	p	p	p
B	mp	mp	mp	p	p	p
C	mp	mp	mp	p	p	p
D	mp	mp	mp	p	p	p

Check 2 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 0.2 M MgCl2
Buffer 0.1 M Tris
Prep. 20% PEG8000

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.	no Zn			1 mM Zn		
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130125

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 50 mM MgCl2
Buffer 0.1 M MES
Prep. 20% PEG6000

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.
Cell:

	1	2	3	4	5	6	
A	5	5.5	6	6.5	7	7.5 + Mg	
B	5	5.5	6	6.5	7	7.5 + Mg	+Zn
C	5	5.5	6	6.5	7	7.5	
D	5	5.5	6	6.5	7	7.5	+Zn

Check 1 20130125

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 0.2 M MgCl2
Buffer 0.1 M Tris
Prep. 36% PEG3350

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.	no Zn			1 mM Zn		
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130125

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 0.2 M CaCl
Buffer 0.1 M Tris
Prep. 45% 2-Methyl-2,4 pentadiol

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.	no Zn			1 mM Zn		
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130125

	1	2	3	4	5	6
A	p	mp	p	mp	mp	p
B	p	p	c	c	mp	c
C	mp	c	mp	mp	c	c
D	mp	c	p	c	c	p

Check 2 201202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	c	p
D	p	mp	p	p	c	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	c	p
D	p	p	p	p	c	p

Check 4 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	c	p
D	p	p	p	p	c	p

Check 5 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	mp	p
D	p	p	p	p	mp	p

Check 6 20130409

	1	2	3	4	5	6

A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	mp	p
D	p	p	p	p	mp	p

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 50 mM CaCl
Buffer 0.1 M Tris
Prep. 20% 2-Methyl-2,4 pentadiol

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.	no Zn			1 mM Zn		
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1

	1	2	3	4	5	6	Destroyed!
A	c	p	mp	c	p	c	
B	p	mp	mp	p	mmp	c	
C	p	mp	mp	p	mp	mp	
D	mp	mp	mp	mp	mp	mp	

Check 2 20130202

	1	2	3	4	5	6
A	p	p	X	X	X	?
B	X	p	p	p	X	mp
C	p	p	p	p	p	X
D	X	X	p	X	X	X

Check 3 20130219

	1	2	3	4	5	6
A	p	p	X	X	X	X
B	X	p	p	p	X	X
C	p	p	p	p	p	X
D	X	X	p	X	X	X

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130124
Protein Vif-EloBC-CBFb
[con] 6 mg/ml **Method** BioRed

Salt 0.25 M NaCl
Buffer 50 mM Tris
Prep. 30% PEG3350

Drop1 condition 0.75 ul sample 0.75 ul
Drop2 condition 0.75 ul sample 1.5 ul

Cat.	no Zn			1 mM Zn		
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130125

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130202

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130222
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.2 M CaCl
Buffer 0.1 M HEPES
Prep. 40% PEG 400

Drop1 condition 1 ul sample 1 ul
Drop2 condition sample

Prep. 1 mM Zn
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130222
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.1 M MgCl2
Buffer 0.1 M Tris
Prep. 8% PEG 8000

Drop1 condition 1 ul sample 1 ul
Drop2 condition sample

Prep. 1 mM Zn
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130222
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.1 M MgCl2
Buffer 0.1 M Tris
Prep. 8% PEG 8000

Drop1 condition 1 ul sample 1 ul
Drop2 condition sample

Prep. 1 mM Zn
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130222
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 50 mM MgCl2
Buffer 0.1 M MES
Prep. 20% PEG6000

Drop1 condition 1 ul sample 1 ul
Drop2 condition sample

Prep. 10% PEG 6000
Cell:

	1	2	3	4	5	6	
A	5	5.5	6	5	5.5	6	
B	6.5	7	7.5	6.5	7	7.5	
C	5	5.5	6	5	5.5	6	
D	6.5	7	7.5	6.5	7	7.5	1 mM Zn

Check 1 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130222
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.2 M MgCl2
Buffer 0.1 M Tris
Prep. 40% PEG3350

Drop1 condition 1 ul sample 1 ul
Drop2 condition sample

Prep. 1 mM Zn
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130219

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130222
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.2 M CaCl
Buffer 0.1 M Tris
Prep. 45% 2-Methyl-2,4 pentadiol

Drop1 condition 1 ul sample 1 ul
Drop2 condition sample

Prep. 1 mM Zn
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130219

	1	2	3	4	5	6
A	rubbish	rubbish	mp	rubbish	rubbish	p
B	rubbish	rubbish	mp	rubbish	c	rubbish
C	rubbish	mp	p	rubbish	c	mp
D	rubbish	mp	p	rubbish	c	c

Check 2 20130307

	1	2	3	4	5	6
A	rub/c ?	rubbish	p	rub/c ?	rub/c ?	p
B	rub/c ?	rubbish	p	rub/c ?	c	rubbish
C	rubbish	p	p	rub/c ?	p	p
D	rubbish	p	p	rub/c ?	p	c

Check 3 20130314

	1	2	3	4	5	6
A	r/p	r/p	p	rub/c ?	rub/c ?	p
B	r/p	r/p	p	r/p	c	r/p
C	r/p	p	p	rub/c ?	p	p
D	r/p	p	p	r/p	p	c

Check 4 20120409

	1	2	3	4	5	6
A	r/p	r/p	p	r	r	p
B	r/p	r/p	p	r	c	r/p
C	r/p	p	p	r	p	p
D	r/p	p	p	r/p	p	c

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130306
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.2 CaCl2
Buffer 0.1 M HEPES

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Prep.	40% PEG 400			20% PEG 400		
Cell:						
	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130307

	1	2	3	4	5	6
A	s, mp	s, mp	p	s, mp	p	p
B	s, mp	s, mp	p	s, mp	p	p
C	s, mp	p	p	p	p	p
D	s, mp	p	p	p	p	p

Check 2 20130314

	1	2	3	4	5	6
A	s/p	s/p	p	s, mp	p	p
B	s/p	s/p	p	s, mp	p	p
C	s/p	p	p	p	p	p
D	s/p	p	p	p	p	p

Check 3 20130322

	1	2	3	4	5	6
A	s/p	s/p	p	s, mp	p	p
B	s/p	s/p	p	s, p	p	p
C	s/p	p	p	p	p	p
D	s/p	p	p	p	p	p

Check 4 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130306
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt
Buffer 0.1 M Tris

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

, 0.1 M MgCl2 4% PEG 8k 0.2 M CaCl2 45% 2-Me-2,4-Pen
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130307

	1	2	3	4	5	6
A	c	c	c	p	p	p
B	c	c	c	p	p	p
C	c	c	salt?	p	p	p
D	c	c	salt?	p	p	p

Check 2 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	salt?	p	p	p
D	p	p	salt?	p	p	p

Check 3 20130322

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	salt?	p	p	p
D	p	p	salt?	p	p	p

Check 4 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	s	p	p	p
D	p	p	s	p	p	p

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130306
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt 0.2 MgCl2
Buffer 0.1 Tris

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Prep. 20% PEG 8000 10% PEG 8000
Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	salt?
B	p	p	salt?	p	p	salt?
C	p	p	p	salt?	p	p
D	p	p	p	p	p	p

Check 2 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	s/p
B	p	p	salt?	p	p	s/p
C	p	p	p	p, middle?	p	p
D	p	p	p	p	p	p

Check 3 20130322

	1	2	3	4	5	6
A	p	p	p	p	p	s/p
B	p	p	salt?	p	p	s/p
C	p	p	p	p, middle?	p	p
D	p	p	p	p	p	p

Check 4 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	s/p
B	p	p	s/p	p	p	s/p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130306
Protein Vif-EloBC-CBFb
[con] 4 mg/ml **Method** BioRed

Salt
Buffer

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Buffer 50 mM MgCl2 0.1 MES 0.2 M MgCl2 0.1 M Tris
Prep. 20% & 10% PEG6000 20% PEG3350
Cell:

	1	2	3	4	5	6
A	5	5.5	6	4	6	8
B	6.5	7	7.5	4.5	6.5	8.5
C	5	5.5	6	5	7	9
D	6.5	7	7.5	5.5	7.5	9.5

Check 1 20130307

	1	2	3	4	5	6
A	p	p	p	p	p	salt?
B	p	p	p	p	p	salt?
C	p	p	p	p	p	p
D	mp	mp	mp	p	p	salt?

Check 2 20130314

	1	2	3	4	5	6
A	p	p	p	p	p	s/p
B	p	p	p	p	p	s/p
C	p	p	p	p	p	p
D	p	p	p	p	p	salt?

Check 3 20130322

	1	2	3	4	5	6
A	p	p	p	p	p	s/p
B	p	p	p	p	p	s/p
C	p	p	p	p	p	p
D	p	p	p	p	p	s/mp

Check 4 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	s/p
B	p	p	p	p	p	s/p
C	p	p	p	p	p	p
D	p	p	p	p	p	s/mp

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130327
Protein Vif-EloBC-CBFb
[con] 3.6 mg/ml **Method** BioRed

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Salt	0.2 M MgCl2	0.2 MgCl2
Buffer	0.1 M HEPES	0.1 M Tris
Prep.	10% PEG400	10% PEG3350

Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130328

	1	2	3	4	5	6
A	mp	mp	mp	p	p	p
B	mp	mp	mp	p	p	p
C	mp	mp	p	p	p	p
D	mp	mp	p	p	p	p

Check 2 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130327
Protein Vif-EloBC-CBFb
[con] 3.6 mg/ml **Method** BioRed

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Salt	50 mM MgCl2	0.1 M MgCl2
Buffer	0.1 M MES	0.1 M Tris
Prep.	5% PEG6000	4% PEG8000

Cell:

	1	2	3	4	5	6
A	5	5.5	6	4	6	8
B	6.5	7	7.5	4.5	6.5	8.5
C	5 (Zn+)	5.5	6	5	7	9
D	6.5	7	7.5	5.5	7.5	9.5

Check 1 20130328

	1	2	3	4	5	6
A	p	p	p	mp	mp	mp
B	p	mp	mp	mp	mp	mp
C	p	p	p	mp	mp	p
D	p	p	p	mp	mp	p

Check 2 20130409

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130327
Protein Vif-EloBC-CBFb
[con] 3.6 mg/ml **Method** BioRed

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Salt	0.2 M CaCl2	0.2 M CaCl2
Buffer	0.1 M Tris	0.1 M Tris
Prep.	22.5% 2-Methyl-2,4 pentantal	12% 2-Methyl-2,4 pentantal

Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130328

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130409

	1	2	3	4	5	6
A						
B						
C						
D						

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130327
Protein Vif-EloBC-CBFb
[con] 3.6 mg/ml **Method** BioRed

Drop1 condition 5 ul sample 5 ul
Drop2 condition sample

Salt	0.25 M NaCl			0.25 M NaCl + 1 mM ZnCl2		
Buffer	50 mM Tris			50 mM Tris		
Prep.	10% PEG3350			10% PEG3350		

Cell:

	1	2	3	4	5	6
A	4	6	8	4	6	8
B	4.5	6.5	8.5	4.5	6.5	8.5
C	5	7	9	5	7	9
D	5.5	7.5	9.5	5.5	7.5	9.5

Check 1 20130328

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C	p	p	p	p	p	p
D	p	p	p	p	p	p

Check 2 20130409

	1	2	3	4	5	6
A						
B						
C						
D						

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date 20130327
Protein Vif-EloBC-CBFb
[con] 3.6 mg/ml **Method** BioRed

Drop1 condition sample
Drop2 condition sample

Salt 0.2 M CaCl2
Buffer 0.1 M HEPES
Prep. 10% PEG400

Cell:

	1	2	3	4	5	6
A	4	4.5	5	5.5	6	6.5
B	7	7.5	8	8.5	9	9.5
C						
D						

Check 1 20130328

	1	2	3	4	5	6
A	p	p	p	p	p	p
B	p	p	p	p	p	p
C						
D						

Check 2 20130409

	1	2	3	4	5	6
A						
B						
C						
D						

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						

Date
Protein Vif-EloBC-CBFb
[con] mg/ml **Method** BioRed

Drop1 condition sample
Drop2 condition sample

Salt
Buffer
Prep.

Cell:

	1	2	3	4	5	6
A						
B						
C						
D						

Check 1

	1	2	3	4	5	6
A						
B						
C						
D						

Check 2

	1	2	3	4	5	6
A						
B						
C						
D						

Check 3

	1	2	3	4	5	6
A						
B						
C						
D						

Check 4

	1	2	3	4	5	6
A						
B						
C						
D						

Check 5

	1	2	3	4	5	6
A						
B						
C						
D						